

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2002, 05:37:35 ; Search time 5265 seconds  
(without alignments)  
16995.607 Million cell updates/sec

Title: US-09-730-559B-1  
Perfect score: 4276  
Sequence: I ttctaccgtttttccctgc.....atcagaaaaaaaaa 4276

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg0.inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

1	3690.4	86.3	167659	2	AC023404	Homo sapi	
c	2	3690.4	86.3	175793	2	AC093300	Homo sapi
3	3690.4	86.3	180130	2	AC093275	Homo sapi	
c	4	2480.4	58.0	157320	2	AC027609	Homo sapi
5	912.6	21.3	6380	9	AB051460	Homo sapi	
6	467.4	10.9	212601	2	AL662793	Mus muscu	
c	7	467.4	10.9	249657	2	AL669946	Mus muscu
8	430	10.1	5715	9	AB023157	Homo sapi	
9	336.4	9.3	71171	9	AC100917	Mus muscu	
10	304.6	7.1	3114	3	AY052043	Drosophil	
c	11	304.6	7.1	103054	3	AC014803	Drosophil
c	12	304.6	7.1	181720	3	AC010114	Drosophil
c	13	304.6	7.1	287018	3	AE003553	Drosophil
c	14	240.8	5.6	205307	2	AC009968	Homo sapi
c	15	238	5.6	97037	9	AC004973	Homo sapi
c	16	238	5.6	126000	9	AF000744	Homo sapi
c	17	238	5.6	135038	9	HUMYXWD703	
c	18	238	5.6	190112	2	AP002354	Homo sapi
c	19	237	5.5	113687	9	AC069281	Homo sapi
c	20	233.8	5.5	179245	9	HUAC002492	Human chr
c	21	231.4	5.4	203959	2	AC064827	Homo sapi
c	22	230.4	5.4	133475	9	AC006254	Homo sapi
c	23	230.2	5.4	148300	2	AC091489	Homo sapi
c	24	230.2	5.4	181532	9	CNS01DW2	Human chr
c	25	230.2	5.4	187696	2	AC008785	Homo sapi
c	26	229.4	5.4	155691	9	AC016868	Homo sapi
c	27	229.4	5.4	214872	2	AC009720	Homo sapi
c	28	228.6	5.3	166889	2	AC027008	Homo sapi
c	29	228.6	5.3	167412	2	AL356352	Homo sapi
c	30	228.6	5.3	169963	9	AL158207	Human DNA
c	31	228.6	5.3	241288	2	AC011499	Homo sapi
c	32	228.4	5.3	147224	2	AP003779	Homo sapi
c	33	228.4	5.3	161262	2	AC087464	Homo sapi
c	34	228.4	5.3	167998	9	AL353662	Human DNA
c	35	228.4	5.3	171265	9	AC092573	Homo sapi
c	36	228.4	5.3	208729	2	AC011667	Homo sapi
c	37	228.4	5.3	215177	2	AC012052	Homo sapi
c	38	228	5.3	162589	2	AC022076	Homo sapi
c	39	228	5.3	180465	9	AC074011	Homo sapi
c	40	228	5.3	196769	2	AC092164	Homo sapi
c	41	228	5.3	207922	2	AC093164	Homo sapi
c	42	227.8	5.3	84388	9	AL589984	Human DNA
c	43	227.4	5.3	159634	2	AC010193	Homo sapi
c	44	227.2	5.3	163891	9	AC092647	Homo sapi
c	45	227	5.3	172789	2	AC104020	Homo sapi

ALIGNMENTS

RESULT	1	AC023404	167659 bp	DNA	linear	HTG 01-MAR-2000
AC023404	1	Homo sapiens clone RP11-619L12, WORKING DRAFT SEQUENCE, 10				
LOCUS		AC023404				
DEFINITION		unordered pieces.				
ACCESSION		AC023404.2	GI:7139760			
VERSION		HTG: HTGS_PHASE1, HTGS_DRAFT.				
KEYWORDS		human.				
SOURCE		Homo sapiens				
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE		1 (bases 1 to 167659)				
AUTHORS		Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N., Anderson S., Baldwin J., Barna N., Bada F., Bozulskiy L., Boukhgalter B., Brown A., Burkett G., Campopiano A., Castle A., Chospel Y., Colangelo M., Collins S., Collipalmo A., Cooke P., DeArrellano K., Deward K., Domino M., Doyle M., Fenestor J., Ferreira P., FitzHugh W., Forrest C., Gage D.,				
JOURNAL		Unpublished				
REFERENCE		2 (bases 1 to 167659)				
AUTHORS		Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N., Anderson S., Baldwin J., Barna N., Bada F., Bozulskiy L., Boukhgalter B., Brown A., Burkett G., Campopiano A., Castle A., Chospel Y., Colangelo M., Collins S., Collipalmo A., Cooke P., DeArrellano K., Deward K., Domino M., Doyle M., Fenestor J., Ferreira P., FitzHugh W., Forrest C., Gage D.,				







```
* 107564 126763: contig of 19200 bp in length
* 126764 126863: gap of unknown length
* 126864 151855: contig of 24992 bp in length
* 151856 151955: gap of unknown length
* 151956 175793: contig of 23838 bp in length.
FEATURES             Location/Qualifiers
     source            1..175793
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="rp11-560A7"
                     /clone_11b="RPCI human BAC library 11"
BASE COUNT      51630 a 36283 c 36000 g 50038 t 1842 others
ORIGIN

Query Match      86.3%; Score 3690.4; DB 2; Length 175793;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 3732; Conservative 0; Mismatches 11; Indels 5; Gaps 3;

Qy 530 gtggagcttgcgagtggaatcggtatcggtatcgaggtgtgtgctacgctggattgat 589
|||||
Db 80480 GTGAGAGCTGGCATCAATTAATGATCGGCTATATGAGAGGTGTGTGCTACGCTGGGATGAT 80421
|||||

Qy 590 accgacctgagctaaatacccaaaaggagctgggagagttgcttctctaatcaacag 649
|||||
Db 80420 ACCGACCTGAGCTAAAATACCCAAAAGGAGCTGGGAGAGTTGGTTCCTCTATCAACAG 80361
|||||

Qy 650 agttacatagctgctatcagtgcccgcttggctcagctgcagcatggagagatagataaa 709
|||||
Db 80360 AGTTACATAGCTGCTATACGTGCCGCTTTGTTTCAGCTGCAGCATGGAGATAGATAAA 80301
|||||

Qy 710 cgggtaagccttactacattttgaaaaattctagaattggtccctcctaattgtgatt 769
|||||
Db 80300 CGGTAAGCCTTATACATACATTTGGAAAATCTTAGAAAATGGTCCCTAAATGTGTGATT 80241
|||||

Qy 770 accaatattagaacgggagcattttatgacaataaagtgcagctgacacatttgcctat 829
|||||
Db 80240 ACCAAATTAGAACGGGAGCATTTATGACAAATAAAGTGACAGCTGACAAATTTGGCTAT 80181
|||||

Qy 830 agagttaatatgctctataacatgaaataatgtcctatgaattcttttatcttca 889
|||||
Db 80180 AGAGTTAATTATGTGCTATATATACATGAAATAATGCTCCTATGAATTTCTTTTATCTTCA 80121
|||||

Qy 890 gtttttgagtagcctaatacagacactacaatttacttgagtttaatttaattctctta 949
|||||
Db 80120 GTTTTTGAGTAGGCTTAATCAGAACACTACAATTTACTTGAGTTAAATTTAATCTTCTCTA 80061
|||||

Qy 950 acttccattcaatctcaatccatccgtccattcattcaacttagttgtgaagtcatcaat 1009
|||||
Db 80060 ACTTCCATTCAATCTCAATCCATCCGTCCTCATTCATTCACATTGTTGTAAGTCATTCAAT 80001
|||||

Qy 1010 aaatattactgaatcccttgttctgtgtgtatatacgaagtatacaaacaggaaatgccttg 1069
|||||
Db 80000 AAATATTTACTGAATCCTTTGTTCTGTGTATATCAAGTATACAAAACAGGAATGCCCTTG 79941
|||||

Qy 1070 aggttctccctctttttttgtttttttttaaactcctggacatagggaagacctcagc 1129
|||||
Db 79940 AGGTTTCTGCCCCCTTTTTTTGTTGTTTTTTAATCTGGGACATAGGGAAGACCTCAGC 79881
|||||

Qy 1130 aagccctatttccaatgaattgtactcacagattttctttttttttttttttttttttt 1189
|||||
Db 79880 AAGCCCTATTCTCAATGAATTGTACTACAGATTTC--TTTTTTTTTTTTTTCTTTTT 79824
|||||

Qy 1190 ccacagccgcacactctcaacgatttattccttagcttgggtgttctcattcaacaaa 1249
|||||
Db 79823 CCACAGCGGCCACCTCTCACCGATTATTCCTTAGCTTGGTGTTCATGATTAACAACAA 79764
|||||

Qy 1250 ogtttttagtcttagggcaagaagtctcctcctcatgattttatttccctagcagataga 1309
|||||
Db 79763 CGTTTTAGTGCTTAGGGCAAGAAGTTCCTGTCTCATGAGTTTTATTTCCTAGCAGATAGA 79704
|||||

Qy 1310 actgtataccttgcagtactactcagagtggtgacctgtgacctccagtcctgttaa 1369
|||||
```

```
|||||
Db 79703 ACTGTATCACTTGCCAGTACTACTCAGAGTGTGGCTGTGGACTGACCTCCAGTCTGTAA 79644
|||||
Qy 1370 acttagtttgtagtagtagagaatttagaccagaatgtgtaataaccacacatactctgg 1429
|||||
Db 79643 ACTTAGTTGTAGTAGAGATAGGAATTTAGACCAGAATGTGTAATCAACACATTACTGGG 79584
|||||
Qy 1430 cacaatgttggtccagctggcgatttttttttcatagaagaccctttattgatagggaa 1489
|||||
Db 79583 CACAATGTTGGTCCAGCTGGCGATTTTTTTTCATAGAAAGCCTTTATTGATGAGGGAA 79524
|||||
Qy 1490 gcaatatattgatattatatttgggggtcaccttttttatttoatggcacactggcaatttc 1549
|||||
Db 79523 GCAATATATTGATTTATATTTTGGGGTCACCTTTTATTTCATGCGACACTGGCACTTTC 79464
|||||
Qy 1550 atgcatgctgaotttgatataccatcacctcagggcatgtgctaaaaatagattgatttta 1609
|||||
Db 79463 ATGCATGCTGACTTTTGATATCCATCCTTGAGGCATTTGTCTAAAATAGATTTGATTTA 79404
|||||
Qy 1610 tcgtgtgttctcaattcaagaatgtaaaaatcatcaagtcagtcagtcagcttttttgcctttt 1669
|||||
Db 79403 TCGTGTGTTCTCAATTCAGATGTAAAAATCATCAAGTCAGTCAGTCAGTTCCTTTT 79344
|||||
Qy 1670 atgtttcatgtcatgtacagtcctactcactggcagtagtaaaaaatttaagatatggtggtg 1729
|||||
Db 79343 ATGTTTCATGTCTATGACAGTCTACTTCACTGGCAGTAAAAAAATTTAAGATAGTGGTGG 79284
|||||
Qy 1730 tcatctacaaactgtaactatattaaagagaaaagatctgtctattcttaagcatggg 1789
|||||
Db 79283 TCATCCTACAAAATCTGTAATCTATTAAAGAGAAAAGTATCTGTTCTATTCTTAAGCATGGG 79224
|||||
Qy 1790 ggaggagacaagattagtagttaacatgctactcttgtttgtttgagatgagatctctct 1849
|||||
Db 79223 GGAGGACAAAGATTAGTATGTTTAACATGCCCTACTTTGTTTGTGAGATGGAGTCTCTCT 79164
|||||
Qy 1850 ccggtcaaccaggctggagtgagtggtgtacagtcctcagctcactccaacctctgcctcccg 1909
|||||
Db 79163 CCGTCACCCAGGCTGGAGTGAGTGGTACAGTCTCAGCTCACTCCAACTCTGCCTCCCG 79104
|||||
Qy 1910 ggttcaagtgtatctcctgcttagcctcccgagtagtggtggaattacaggcatataccac 1969
|||||
Db 79103 GGTTCAAAGTGATTTCTCCTGCCCTTAGCCTCCCGAGTAGTGGAAATTTACAGGCATATACCAC 79044
|||||
Qy 1970 catgccacaacaattgtgtatttttagtgagacagggtttcacggtgtggtcagcgc 2029
|||||
Db 79043 CATGCCCAACAATGTTGTATTTTTTAGTGGACAGGGTTTCACCGTGTGGTCAGGCC 78984
|||||
Qy 2030 agtttcaaaactcctgacctcaagggtatccaacctgcctcaccctccaaagtgtcgggatt 2089
|||||
Db 78983 AGTTTCAAACTCCTGAGCTCAAGGATCCACCTGCCTCACCCCTCAAGTGTCTGGGATT 78924
|||||
Qy 2090 acaggcatgagcacccaccatgctggcctacttggttttttatgacacactaaaaata 2149
|||||
Db 78923 ACAGGCATGAGCCACCCACCATGCTGGCTACTTGGTTTTTTTATGCACTAAAAAATA 78864
|||||
Qy 2150 cctacatcctcagctctattccaacataagtttcagagctgtggattggtcattagaa 2209
|||||
Db 78863 CTACATCTCACTGCTTTATTCACATAAAGTTTCAGACTGTGGATGGTCATTAGAA 78804
|||||
Qy 2210 attcagactgaattgtgttctcctctgcaatgaaatccttggccagtggttcactgcctc 2269
|||||
Db 78803 ATTACAGACTGAATTTGTGTGTTCTCTGCAATGAATCCCTTTGCCAGTGTTCATGTCACTC 78744
|||||
Qy 2270 tgtagacattatggagcagcctagaggccagaagccagtgctctccttatgacctgctct 2329
|||||
Db 78743 TGTAGACATTATGGAGACGCTTAGAGCCAGAGCCAGTGTCTCTCTTATGCCCTGCCTCT 78684
|||||
Qy 2330 tctcggctctcagcacctcttctctcttcttcttcttcttcttcttcttcttcttcttct 2389
|||||
Db 78683 TCTTGGGCTCTGTGACACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 78624
|||||
Qy 2390 ttttttagaggagggtctcactctgtcaccacaggtgtcaccagctggagacagatcacatcag 2449
|||||
```

Db	78623	TTTTTTTAGAGGGAGGGTCTCACTCTCTCA	CCCCAGGCTGGAGCAGCAAAATCACAAATCATG	78566
Qy	2450	actcaactgcattctctctctctctctctctt	gtgttcacgtgtaactctgtgcaggtatctctgtcag	2509
Db	78563	ACTCACTGCATGTTCTTCTCCTTTTGTGTCAT	TGGCTAAATCTTGGTCAGCAATTCCTTGTGTCAG	78504
Qy	2510	agctgggtggcacacagtgctgtgtgaacagctg	ctgtgaaggagtttcagccatgaattctc	2569
Db	78503	AGCTGGGTGGCACCAGTGCCTGGTGACAGCTGC	TGTGAAGGGAGTTTCAGGCATGAATCTC	78444
Qy	2570	tccagacttaaaaaataaccagctctcttcttagct	gatgaattataaaccagagtgactgtta	2629
Db	78443	TCCAGACTAAAAATAACCAGCTCTTTTCTAGCT	GATGAATTAATAAACACAGTGACTCTTA	78384
Qy	2630	atgcttgaaaggttccaatgacaggttgcccgatag	aacgcctggaaacagggcccaagtttta	2689
Db	78383	ATGCTTGAAAGGTTTCACATGACAGGTTGGCGAT	TAGAACGCTGGGAACAGGCCAGTTTAA	78324
Qy	2690	gaattcaacctctgaacttttagactcagggtgaac	catcttactagagaagaacaaagca	2749
Db	78323	GAATTCACCTCTGACATTTAGACTCAGGTGAAC	CATTCTTACTGAGAAAGAACAAAGCA	78264
Qy	2750	gggttttagactgtgaatccctatggctgcctc	-tttttttttttttttaacagagttccag	2808
Db	78263	GGGTTTAGACTGTGAATCCCTATGGCTGCATCT	TTTTTTTTTTTTTTTTTTTAAACAGAGTCCAG	78204
Qy	2809	gtttgtaattataaccaacaactgtgcacctataata	tagaaaccacacagccagcgtctttt	2868
Db	78203	GTTTGTGATTATACCCAAACATGTGTACATPAT	TAATATAGAAACCCAGCAGCCGCTTTT	78144
Qy	2869	acgacagctcagaatctgtgacgcagtagtcaggc	atcttcacacgcagcttgaaattg	2928
Db	78143	ACGACAGCTCAGAAATCTTGTGACGCGTAGTGC	AGGCATCTTCACACCAACTTGAATATTG	78084
Qy	2929	aagtgcattgtgtgaacctggatctcttagttgatt	gtttgttaattatgattccac	2988
Db	78083	AAGTGCAGTGTGTGGAACTTGGATCACTTGTAG	TGTGTGTTTGTAAATATGATTTCCAC	78024
Qy	2989	atatgacaaaaatccagatccactcaataataaat	gagggtttatgtctcatgaaataactcc	3048
Db	78023	ATATGACAAAAATCCAGATCCACTAATTAATAA	TGAGGGTTTATGCTATGAATAATCTCC	77964
Qy	3049	tgtgggttaattcctaataacttagtctctaaacag	tgtggcttcacttcatagtctgc	3108
Db	77963	TGTGGGTTTAAATCTCATAACTTCTAGTCTAAAC	AGATTTGGCTTCACATCATGATCTGTC	77904
Qy	3109	tcaaatccttttctcttaaaggatgtttatttaata	agaagaaaaaaatgtaaaatgatag	3168
Db	77903	TCAAATCTCTTTTCCCTTTAAGGATGCTTTTAT	TTAATAAG-AAAAAATCTAAATGATAG	77845
Qy	3169	ataataaagccttactaggttcttaaagaatgaact	atccatatatttcagtaaatgaata	3228
Db	77844	ATAATAAAGCCTTACTAGTCTCTTAAAGATGAAC	TATCCATATTTTCAGTAAATGAATA	77785
Qy	3229	attagctcctctcttttgggcaacctggaaacagatt	ccaattcagatagtggggtgaaatgt	3288
Db	77784	ATTAGTCTCTCTCTTTGGGCACCTTGGAACAGAT	TCATTCAGATAGTGGGTGGAAATGT	77725
Qy	3289	acatgatggttaagcattgtctggcctagtcactgaa	aaatgtaaaactctatttttgatt	3348
Db	77724	ACATGATGGTAAGCATTTGCTGGCCTACTCACTG	AAAAATGTAAACTCTTATTTTGTATT	77665
Qy	3349	gcaggtggaagttaagccatatgtcttcttgatgatc	agctgtgtgatgaatgtccaggggggc	3408
Db	77664	GCAGGTGGAAGCTTAAGCCATATGCTTGGATGAT	GCAGCTGTGTGATGAATGTCAAGGGGC	77605
Qy	3409	ccgttgtggggggaattgtcccaattttctgtgcta	atgttcaactgtctgcagtatta	3468
Db	77604	CCGTTGTGGGGGAATTTGCTCCCAITTTTCTGT	CTCTAATGTTTCTCTGTCAGTATTA	77545
Qy	3469	ctgtgaaattatgtcggctgctctatccattctcgt	ctggcaggggaattcccaagccct	3528
Db	77544	CTGTGAATATTCCTGGGTGCTATCCATTTCTGT	CTGCTGGCAGGGAATTCACAAGCCCT	77485

QY	3529	ggtgaagaagcgggtgacgcgcctcgccatatttcattccgtcggaactaagataac	3588
Db	77484	GGTGAAGGAAGCGGTGACGCCCTCGGCATATTTTCATTCGGCTGGAATTAAGGATAAC	77425
QY	3589	tgcagtgctcatttttcagggcctcagaataaagtcgaactctctctgttcattctgaccccttc	3648
Db	77424	TGCAGTGCCTCAITTTTCAGSCCTCAGAATAAGTGCACCTCTCTGCTTCATTCGACCCCTTC	77365
QY	3649	ctcaacctcttcacgcctggcgcgtgcctcttttgtagcagctgtaacttaactatagataaa	3708
Db	77364	CTCAACCTCTTCACGCTGGCATGTCCTTTGTGTAGCAGCTGTAACTTAATAGTATATAA	77305
QY	3709	tgaagaagatgacctataatagtggttttgtagatctcttgctcactgcgaacaatat	3768
Db	77304	TGAAGAAGATGACCTATTAATATAGGTGTTTGTGTAGATCTTGTGTCACTGCAACAATAT	77245
QY	3769	gaactccttttctgattgacctcgggtgtgcgatggaagtttattctctgtttgtctgctgg	3828
Db	77244	GAACTCCTTTTTCGTATTGTCATCGGTTGCATCGGAAGTTTATCTCTGTGTTTCTGCTGG	77185
QY	3829	aaacaagagatccaaactctcgtcaacatttcttagagagagagagaaatatataa	3888
Db	77184	AAACCAAGAGGATCCAAACTTCCTGCACATTTTCTTAGAGGAGAGAGAAATATATAA	77125
QY	3889	agagaatgaacaatatagatattttggggttttaattaaattattgtaataataaac	3948
Db	77124	AGAGAATGAACAATAGAGTATTTTGGGTTTTTAAATTAATATTGTTAAATAATAAC	77065
QY	3949	atataagaatacttttattataaataaccatgcacaataaacaactatcggtctatctgcaca	4008
Db	77064	ATATAGAATACTTTTATTAATAAATCAACCATGCACAATATACACTATCGGTCTATCTGACA	77005
QY	4009	gttttcccccgagggaagtgtcttggcttttcccttctcttttttttttttttttttttttt	4068
Db	77004	GTTTTCCCCCAGGGAAGTGCTTTTGCCCTTTTCCCTTTTCTTTTCTTTTCTTTTCTATCTT	76945
QY	4069	tttgtctctctcttttttttccatcccttttaatttttttaaacagaagaaggaagttca	4128
Db	76944	TTTGTGCTCTCTCTTTTTCATCCCTTTTAAATTTTTTAAACGCAATGGAGGAAGTTTA	76885
QY	4129	acaattttaatgaaagacgttagagcaacaataagcataagcaagactgagcagc	4188
Db	76884	ACAAATTTTAATGGAAGACATGTTAGAGCAACAACAATGCATAAAGCAAGACTGACGAGC	76825
QY	4189	attataattaatttcagggttttgaggctgaacataatttcattatccctcaaaaagtt	4248
Db	76824	ATTATAATTAATTTTCAGGGTTTTCAGGCTGAACATAATTTTCATTATCCCTCAAAAAGTT	76765
QY	4249	acaccacatcagaaaaaataaaaaa 4276	
Db	76764	ACCACCACATCAGAAAAATAAAAAA 76737	
RESULT	3		
AC093275			
LOCUS		180130 bp	DNA linear HTG 16-AUG-2013
DEFINITION		Homo sapiens chromosome 5 clone RP11-379A13, WORKING DRAFT	
SEQUENCE		7 unordered pieces.	
AC093275			
VERSION		AC093275.1	GI:15193409
KEYWORDS		HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEPIN.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		1 (bases 1 to 180130)	
TITLE		DOE Joint Genome Institute.	
JOURNAL		Sequencing of Human Chromosome 5	
REFERENCE		Unpublished	
AUTHORS		2 (bases 1 to 180130)	
TITLE		DOE Joint Genome Institute.	
REFERENCE		Direct Submission	

JOURNAL

Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----

Project Information

Center Project Name: 551914  
Center clone name: RPCI-11\_379A13  
-----

Summary Statistics

Consensus quality: 171964 bases at least Q40  
Consensus quality: 176669 bases at least Q30  
Consensus quality: 177551 bases at least Q20  
Estimated insert size: 192980; agarose-fp estimation  
Estimated insert size: 179530; sum-of-contigs estimation  
Quality coverage: 10.75 in Q20 bases; agarose-fp estimation  
Quality coverage: 10.75 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1139: contig of 1139 bp in length  
\* 1140 1239: gap of unknown length  
\* 1240 4028: contig of 2789 bp in length  
\* 4029 4128: gap of unknown length  
\* 4129 9194: contig of 5066 bp in length  
\* 9195 9294: gap of unknown length  
\* 9295 27293: contig of 17999 bp in length  
\* 27294 27394: gap of unknown length  
\* 27394 45640: contig of 18247 bp in length  
\* 45641 45740: gap of unknown length  
\* 45741 79464: contig of 33724 bp in length  
\* 79465 79564: gap of unknown length  
\* 79565 180130: contig of 100566 bp in length.

FEATURES

source

1. 180130  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-379A13"  
/clone\_lib="RPCI human BAC library 11"  
52554 a 37538 c 37332 g 52080 t 626 others

BASE COUNT

ORIGIN

Query Match 86.3%; Score 3690.4; DB 2; Length 180130;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 3732; Conservative 0; Mismatches 11; Indels 5; Gaps 3;  
  
QY 530 gtggagcttgcatgtaagctacggtatcggtatcgaggtgtgtgtctacgctgggattgat 589  
Db 15646 GTGGAGCTTGCGATGATATGATCGCTATATGAGGCTGTGTCTACGCTGGGATTGAT 15705  
QY 590 acgacccctgagctaaataaccacaaaggagctgggagagttgcgtctcttaatacaacag 649  
Db 15706 ACCGACCCCTGAGCTAAATACCAAAAGGAGCTGGGAGATTGCGTTCTCTAATCAACAG 15765  
QY 650 agttacatagctgctatcagtcagccgctttgttcagctgcagcatggagagatagataaaa 709  
Db 15766 AGTTACATAGCTGCTATCATAGTCCCGCTTTGTTTCAGCTGCAGCATGGAGATAGATAAA 15825  
QY 710 cgggtaagccttatactacattttggaaaaattctagaattggtcctctaaatgtgtgatt 769  
Db 15826 CGGGTAAGCCTTATACTACATTTTGGAAAAATCTAGAAATGGTCTCTTAATGTGTGATT 15885  
QY 770 accaatattagaacgggagcattttatgacataaagtgcagcgtgacaaattttgcctat 829  
|||||

Db 15886 ACCAATATTAGAACGGGAGCAATTTATGACAAATAAGTAGCAGCTGACAAATTTTGCCTAT 15945  
QY 830 agagtaaatatgtctataatacatgaataatgtctctatgaattctctttatctcttca 889  
Db 15946 AGAGTTAATTTATGGTCTATATAATACATGAATAATGTCCTATGAATTTCTTTTATCTTTCA 16005  
QY 890 gttttttgagtagcctaatacagaacactacaatttacttgaagttaatttaattctctcta 949  
Db 16006 GTTTTTTGAGTAGCCTAATCAGAACACATACAAATTTACTTGAGTTAATTAATCTCTCTA 16065  
QY 950 acttccattcaatctcaatccatccgtccattcacttagtttgaagtcattcaat 1009  
Db 16066 ACTTCCATTCATCTCAATCCATCCGTTCCATTCATTCATTTAGTTTGAAGTCATTCAT 16125  
QY 1010 aaatattactgaatcccttctgtgtgttatatacaagtatcaaacagagaatgcccttg 1069  
Db 16126 AAATATTACTGAATCCTTTCTGTTGTATATCAAGTATACAAACAGGAATGCCCTTG 16185  
QY 1070 aggttctcgccttt 1129  
Db 16186 AGGTTTCTGCTGCCCTTTTGTGTTTGTGTTTAAATCTGGGACATAGGGAAGACCTCAGC 16245  
QY 1130 aagccatttctcaatgaattgtactcaagatttcttttttttttttttttttttttttttt 1189  
Db 16246 AAGCCCTATTTCTCAATGAATTGTACTCACAGATTTC---TTTTTTTTTTTTTTTCTTTTT 16302  
QY 1190 ccacagcgcacacctctcacgatttattccttagcttggtgtttctatgtattcaacaaa 1249  
Db 16303 CCACAGCGCCACCTCTCACGATTATTCTCTAGCTTGGTGTTCATGTATTCAACAAA 16362  
QY 1250 ogttttagtgcttaggcaagaattcctgtccctcatgagtttatttctcagagataga 1309  
Db 16363 CGTTTAGTGCTTAGGGCAAGAAGTTCCTGTCCTCATGAGTTTATTTCCTAGCAGATAGA 16422  
QY 1310 actgtatcaacttgccagttactactcagagttgtgcccctgtggaactgcctcagctctaa 1369  
Db 16423 ACTGTATCACTTGCCAGTACTACTCAGAGTGTGGCCTGTGGACTGACCTCCAGTCTGTAA 16482  
QY 1370 acttagttgtagtgagtaggaatttagaccagaattgtgtaatacaaccacattactggg 1429  
Db 16483 ACTTAGTTGTAGTGAGATAGGAATTTAGACCAGAAATGTGTAATCAACCCACATTACTTGGG 16542  
QY 1430 cacaatgttggtccagctggcgatttttttttcatagaagccctttatgatgaggaa 1489  
Db 16543 CACAATGTTTGGTCCAGCTGCGGATTTTTTTTCATAGAAGCCCTTTATTGTAGAGGGAA 16602  
QY 1490 gcaatatattgatttatatttttgggggtcaccttttttatttcattgcacactggcacttc 1549  
Db 16603 GCAATATATTGATTATATTTTGGGGTCACTTTTATTTCATGGCACACATGGCACATTTC 16662  
QY 1550 atgcatgctgactttgatataccatcactcaggtcaggtcattgtgctaaataagattgattta 1609  
Db 16663 ATGCATGCTGACTTTGATATCCATCTACTGAGGCATTTGCTAAAAATAGATTGATTTTA 16722  
QY 1610 tctgtgttctcaatcaatcaagatgtaaaaatcatcaagtcagtagcagtttttgcctttt 1669  
Db 16723 TCGTGTGTTGTTCTCAATTCAGATGTAAAAATCATCAAGTCAGTAGCAGTTTTTCGCTTTT 16782  
QY 1670 atgttcatgtcatgtacagcttacttctcactggcagtaaaaaatttaagatagtggtgg 1729  
Db 16783 ATGTTTTCATGTCATGTACAGTCTACTTTCCTGGCAGTAAAAAATTTAAGATAGTGTGG 16842  
QY 1730 tcatctcaaaactgtgaattctattataaagaaagatactgttcttattctaagcatggg 1789  
Db 16843 TCATCCCTACAAACCTGTGAATCTATTAAAGAGAAAAAGTATCTGTTCTATTCTTAAGCATGGG 16902  
QY 1790 ggagggaacaagattagtagtgaacatgcctactttgtttgtttgagatgaggtctctct 1849  
Db 16903 GGAGGGACAGATTAGTATGTTAAATGCCCTACTTTGTTTGTGTTGAGATGGAGTCTCTCT 16962  
QY 1850 ccgtcaacccaggctggagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1909  
Db 16963 CCGTCACCCAGGCTGGAGTGCAGTGGTACAGTCTCAGCTCACCTCCACACCTCTGCCTCCCG 17022







```

Db 19182 TTTTCTCTCTCTTTTTCATCCCTTTTAAATTTTAAACAGCAATGGAGGAATTA 19241
QY 4129 acaatttttaagaaagagcatgttagagcaaaacaataagcaaaagactgagcagc 4188
Db 19242 ACAATTTTAAATGGAAGAGCATGTTAGAGCAAAACAATGATAAGCAAGACTGAGCAGC 19301
QY 4189 attataaatttttcagggttttagcgtgaacataatttcattatccctcaaaaagtt 4248
Db 19302 ATTATAATTAATTTTCAGGGTTTGGCGTGAACATAATTTCATTATCCCTCAAAAAGTT 19361
QY 4249 accaccacatcagaaaaaataaaaaa 4276
Db 19362 ACCACCACATCAGAAAAATAAAAAA 19389

RESULT 4
AC027609/c
LOCUS
DEFINITION Homo sapiens chromosome 5 clone RP11-80E21, WORKING DRAFT SEQUENCE,
34 unordered pieces.
ACCESSION AC027609
VERSION AC027609.3 GI:9958118
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157320)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 157320)
Waterston,R.H.
Direct Submission
Submitted (30-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:7637328.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH080E21
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-primer ET; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 138437 bases at least Q40
Consensus quality: 143665 bases at least Q30
Consensus quality: 145934 bases at least Q20
Insert size: 167000; agarose-fp
Quality coverage: 3.44 in Q20 bases; agarose-fp
Quality coverage: 3.89 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1657: contig of 1657 bp in length
* 1658 1757: gap of unknown length
* 1758 3454: contig of 1697 bp in length
* 3455 3554: gap of unknown length
* 3555 5136: contig of 1582 bp in length

FEATURES
source
1..157320
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-80E21"
1..1657
/note="assembly_name:Contig14"
1758..3454
/note="assembly_name:Contig15"
3555..5136
misc_feature
misc_feature
misc_feature
```



```
|||||
Db 42586 ACACATATAATAGAAACACGAGCGAGCTTTTACGACACCTCAGAAATCTGTGACGCA 42527
QY 2895 gtagcagcgcattctcacaccgacttgaaattgaagtgcagttgtggaacttgatc 2954
Db 42526 GTAGTCAGGCATCTTCACACCAACTTGAATATGAAGTGCAGTGTGTGGAACTTGGATC 42467
QY 2955 atcttagtgatttggtttaattatgatcccaatgatgacaaaaatccagatccactaa 3014
Db 42466 ATCTTAGTGTGATTTTGTAAATATGATATCCATATGACAAAATCCAGATCCACTAA 42407
QY 3015 ttaaatgaggggttatgtctatgaataatctcgtggtggtttaatctcataacatctta 3074
Db 42406 TTAATAATGAGGGTTATGCTATGATAATATCTCTGCGGTTTAATCTCATAAATCTTA 42347
QY 3075 gtctaaacagttgggttcacttcattgatgtctgctcgaatccttttctcttaagatg 3134
Db 42346 GTCTAAACAGTTGGCTTCACATTCATGATGCTGCTCAAAATCCTTTTCTCTTAAAGGATG 42287
QY 3135 ttatttaataagaataaaatgtaaatgatagataataaagccttactagattctta 3194
Db 42286 TTTATTTAATAAG - AAAAAAATGTAATAATGATAGATAATAAAGCCTTACTAGGTCTCTTA 42228
QY 3195 aaagatgaactatccatatcttcagtaaatgaataattagtcctctctcttgggcacctt 3254
Db 42227 AAAGATGAACATATCCATATTTTCAGTAATGAATAATTAGTCTCTCTTTGGGCACCTT 42168
QY 3255 ggaacagattcattcagatagtggtggaatgtacatgtatgtaagcatgtcgtggcct 3314
Db 42167 GGAACAGATTTCATTCAGATAGTGGTGGAAATGTACATGTATGTTAAGCATTTGCTGGCCT 42108
QY 3315 agtctctaaataatgaactcttatttttaattgcaagtggaagttaagccatatgtct 3374
Db 42107 AGTCACTGAAAATGTAAACTCTTATTTTGATTGACGTTGAAGTTAAGCCATATGTCT 42048
QY 3375 tggatgatcagctgtgtgatgaatgtcagggggcccggtgtgggggaaatttgcctcat 3434
Db 42047 TGGATGATCAGCTGTGTGATGAATGTCAGGGGGCCCGTGTGGGGGAAATTTGCTCCAT 41988
QY 3435 tttctgtgctaattcactgctgcagttactgtgaattactgtgaattactgctggcgtgctatcc 3494
Db 41987 TTTTCTGTCTAATGTACCTGTCTGCAGTATTAAGTGTGAATATGCTGGGCTGTATCC 41928
QY 3495 attctgctgagcaggaattccaaagccctgtgtaagaaagcgtgaacccctc 3554
Db 41927 ATTCTGCTGTGCAGGGAATTTCCACAGCCCTGTTGTAAGGAAGCGGTGACTCCCTC 41868
QY 3555 ggcattttcattccgctggaactaaaggataaactgcagtgctcattttcagggcctcaga 3614
Db 41867 GGCATATTTTCATCCGCTGGAATAAAGGATAAATGCAGTGTCTCATTTTCAGGCCCTCAGA 41808
QY 3615 ataagtgcactctctgttcattctgcacccctctcactcactcctcagcgtggcatgtcc 3674
Db 41807 ATAAGTGCATCTCTCTGTTCATTTCTGACCCCTTCCTCAACCTCTTCACGCTGGCATGTCC 41748
QY 3675 tttgtagcagctgtgaacttaactatagtataatgaagaagatcacctataataggt 3734
Db 41747 TTTGTAGCAGCTGTGAACCTTAACATATAGTATAATGAAGAAGTGAACCTTATTAATAGGT 41688
QY 3735 gttttgtagattctgttcactgcaaaaataatgaactccttttctcgtattggcattcgg 3794
Db 41687 GTTTTGTAGATCTCTGTGTCACGTCAACAAATATGAACCTCTTTTTCGTATTGCCATCGG 41628
QY 3795 gttgcagtgaagttttattctctgttctgtggaaccaaagagatccaaactctctgc 3854
Db 41627 GTTGCATGGAAGTTTATTCTCTTGTGTTTGTCTGGAACCAAGAGGATCCAAACTTCTCTGC 41568
QY 3855 aacattttcttagagagagagagaaataattaaagagaaaatgaacaatacagattttt 3914
Db 41567 AACATTTCTTAGAGAGAGAGAGAAATATTAAGAGAGAAATGAACAATAAGAGTATTTT 41508
QY 3915 ggggttttaataaattattgttaataataacatatgaatactttttataaataa 3974
|||||
Db 41507 GGCTTTTAAATTAATTAATTAATATAACATATAAGAAATCTTTTATTAAAAATAA 41448
QY 3975 ccatgcaacaataaactactcgtctatcgtacagtttttccccccagggaagtgttttg 4034
Db 41447 CCATGCAACAATAAACAATATCGTCTATCTGACAGTTTTCCCCCAGGGAAGTGTGTTG 41388
QY 4035 ccttttccctctcttttttttttttttttttttttttttttttttttttttttttttttt 4094
Db 41387 CCTTTCCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 41328
QY 4095 tttttaaatttttaacagcaatggaggaagttaacaatttttaataatttttcagggttttga 4154
Db 41327 TTTTAAATTTTAAACAGCAATGGAGGAAGTTAAACAATTTTAAATGGAAGAGCATGTT 41268
QY 4155 agagcaaaaataatgcataagaacagactgagcagcattataatttaatttttcagggttttga 4214
Db 41267 AGAGCAAAACAATAATGCAATAAGCACTGAGCAGCATATATAATTAATTTTCAGGGGTTTGA 41208
QY 4215 ggctgaacataaattcattatccctcaaaaagtaccaccacatcagaaaaaataaaaaa 4274
Db 41207 GGCTGAACATAATTTTCATTATCCCTCAAAAAGTTACCACCACATCAGAAAAATAAAAA 41148
QY 4275 aa 4276
Db 41147 AA 41146
|||||
RESULT 5
AB051460
LOCUS AB051460 6380 bp mRNA linear PRI 07-FEB-2001
DEFINITION Homo sapiens mRNA for KIAA1673 protein, partial cds.
ACCESSION AB051460
VERSION AB051460.1 GI:12697890
KEYWORDS
SOURCE Homo sapiens cDNA to mRNA, clone:fg00690.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Nagase,T., Kikuno,R., Hattori,A., Kondo,Y., Okumura,K. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes.
XIX. The complete sequences of 100 new cDNA clones from brain which
code for large proteins in vitro
JOURNAL DNA Res. 7 (6), 347-355 (2000)
MEDLINE 21082932
REFERENCE 2 (bases 1 to 6380)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-2000) Osamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
Fax:81-438-52-3914)
FEATURES
source
1. .6380
location=Homo sapiens"
/db_xref="taxon.9606"
/clone="fg00690"
/notes="vector:pBluescriptII SK plus"
69. .2207
<69. .2207
/gene="KIAA1673"
/gene="KIAA1673"
/notes="Start codon is not identified."
/codon_start=1
/product="KIAA1673 protein"
/protein_id="BAB21764.1"
/db_xref="GI:12697891"
/translation="MGDYGFLVQSNLGNKSAFVRFHPHLOPHHONATPSPAF
INNNTAANGSSAGANLFPAPATHNQDEILGSEKAKSQOQEQDPLEKQLSPPQ
EAGLLPETERAKSEENQNGEKIRIESPVLTFDYEATLGLTSTQPLTS
SASLTGFSNWSAIAFPSSTIINEDASFFHOGVPAASANNALLFQNPFPVHSPGF
GGSFSPQIGLSLPHHPHFQHHHSQHQQRRSPASPPPTFHRNAAPNQLPHLAN

```

NLNKPPSPNSVQSPPTPSSWSFGGGYGGWGSQGRHRRGLNGGITPLNISPL  
KNFASNH1QLQKARYAPSSAFAPKSMEDSLNRADNIFPDPDRPTFOMHSLSD  
INRAENDT1KARTYGRGSSSLFPMEDGDLDDGRGDOPHLSGLSPHCFSHONGERV  
ERYSRKVFVGGPLPPDIDEITASFRRSGLPLVDWPHKAESKSYPPPKGYAFLLFQDE  
SSVQALIDACIEDKLYKLVSSPIKPKQVQIRPNWLSDSDFVMDGSQPLDPRKTFI  
RGVPRPLRAVELAMIDKLYGGVYAGTIDPELKYPKGAGRAVAFSNOQSYIAISA  
REVQHQHEDIKRVEKVPYVLDQDCQAGRCGKFPFFCANVTCLQYCYECWA  
AHSRAGREFHRLVKEGDRPHISFRWN"  
BASE COUNT 1895 a 1163 c 1217 g 2105 t  
ORIGIN

Query Match 21.3%; Score 912.6; DB 9; Length 6380;  
Best Local Similarity 99.0%; Pred. No. 6.5e-186;  
Matches 918; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3350 caggtggaagttaagccatagctcttgatgatcagctgtgtgatgaatgtcagggggcc 3409  
|||||  
Db 1977 CGGTTGGAAGTTAAGCCATATGCTTGGATGATCAGCTGTGTGAATGTCAGGGGCC 2036  
|||||  
QY 3410 cgttggggggaaattgtccattttctgtgctaaattgtacctgtctgcagtattac 3469  
|||||  
Db 2037 CGTTGGGGGGAATTTGCTCACTTTTCTGTGCTAATGTTACCTGTCTGCAGTATTAC 2096  
|||||  
QY 3470 tgtgaattatgtggctgctatccattctcgtcgtgaggaattccacaagcccttg 3529  
|||||  
Db 2097 TGTGAATATTGCTGGCTGCTATCCATTCCTGCTGCGAGGAAATCCACAAGCCCTG 2156  
|||||  
QY 3530 gtgaaggaagcgggtgacgcctcggcatatttcctccgtgcgaactaaaggaatac 3589  
|||||  
Db 2157 GTCAAGNAGGCGGTGACGCCCTCGGCATATTTTCTCCGTGGAATTAAGGATAACT 2216  
|||||  
QY 3590 gcaagtctcatcttgcaggaatgaagtcacactctctgttcaattctgacccttcc 3649  
|||||  
Db 2217 GCAGTGTCTCATTTTCAGGCTCAGATAAGTGCACCTCTTCTGCTTCTGACCCCTTCC 2276  
|||||  
QY 3650 tcaacctctcagctggcatgctctttgttagcagctgtgaacttaactatagataat 3709  
|||||  
Db 2277 TCAACCTCTTCAGCGTGGCATGCTCTTTGTAGCAGTCTGTAACTTAATAGTATATAT 2336  
|||||  
QY 3710 gaaaagaatgacctataatagtggtttgttagatctctgtctcactgcaacaataatg 3769  
|||||  
Db 2337 GAAAGAATGACCTATAATAGGTGTTTGTAGATTCTGTGTCACGCAACCAATATG 2396  
|||||  
QY 3770 aactccttttgcattgcatcgggtgtcaggaattttatctctgtttgtctgga 3829  
|||||  
Db 2397 AACTCTCTTTTCTGATGTCATCGGTTGCATGGAAGTTTATCTCTGTTTGTCTGGA 2456  
|||||  
QY 3830 aaccaagagatccaaactctcgaactttcttagagagagagagaaataataaa 3889  
|||||  
Db 2457 ACCAAGAGATCCAAACTCTCTGCAACATTTCTTAGAGGAGAGAGAAATATAAA 2516  
|||||  
QY 3890 gagaatgaaacaatagagattttgggttttaataataattatgttaataataaaca 3949  
|||||  
Db 2517 GAGAAATGAACAATAGAGATTATTTGGGTTTTTAATTAATATTGTTTAATAATATA 2576  
|||||  
QY 3950 tataagaatacttttataaataaccatgcgaacataaacactatcggtctatctgcag 4009  
|||||  
Db 2577 TATAGAATACATTTTATTAAAAATACCATGCAACAATAACACTATCGGTCTATCTGAC 2636  
|||||  
QY 4010 ttttccccagggaagtgtcttttgcctttctcttcttttttttttttttttttttttt 4069  
|||||  
Db 2637 TTTTCCCCCAGGGAAGTGTCTTGGCTTTTCCCTTTCTTTCTTTTCTTTTCTTTTCT 2696  
|||||  
QY 4070 ttgtctctctcttttttccatcccttttttaattttttaacagcaatggaggaagttaa 4129  
|||||  
Db 2697 TTTGTCTCTCTCTTTTCTCATCCCTTTTAAATTTTAAAGCAATGGAGGAAGTTAA 2756  
|||||  
QY 4130 caatttttaataagagcatgttagagcaaaaataataagcaagactgagcagca 4189  
|||||  
Db 2757 CAATTTTAAATGGAAGAGCATGTTAGAGCAAAATAATGCAATAAGCAAGACTGAGCA 2816  
|||||  
QY 4190 ttataattaattttcagggtttttgaggtgcaacataatttcatctatccctcaaaagtta 4249  
|||||

Db 2817 TTATAATTAATTTTCAGGTTTTCAGGCTGACATAATTTTCATTTATCCCTCAAAAGTTA 2876  
|||||  
QY 4250 ccaccacatcgaaaaaataaaaaaa 4276  
|||||  
Db 2877 CCACCACATCAAAAAATAAAAAA 2903  
|||||

RESULT 6  
AL662793 212601 bp DNA linear HTG 15-DEC-2001  
LOCUS Mus musculus chromosome 11 clone RP23-186D6, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, in unordered pieces.  
ACCESSION AL662793  
VERSION 2  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (sites)  
REFERENCE Direct Submission  
AUTHORS Submitted (14-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,  
TITLE Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
JOURNAL humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
COMMENT On Dec 18, 2001 this sequence version replaced gi:17736659.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BM186D6  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 211960 bases at least Q40  
Consensus quality: 212056 bases at least Q30  
Consensus quality: 212130 bases at least Q20  
Insert size: 212301; sum-of-contigs  
Insert size: 209170; 13.0% error; agarose-fp  
Quality coverage: 13.41x in Q20 bases; sum-of-contigs Quality  
coverage: 14.00x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES  
source Location/Qualifiers  
1..212601  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/clone="RP23-186D6"  
/clone\_lib="RPCI-23"  
1..36832  
/note="assembly\_fragment:05552  
fragment\_chain:1  
clone\_end:SP6  
vector\_side:left"  
36933..51344  
/note="assembly\_fragment:03482  
fragment\_chain:1"  
51445..201162  
/note="assembly\_fragment:04981  
fragment\_chain:1"  
201263..212601  
/note="assembly\_fragment:04110"  
BASE COUNT 59008 a 45779 c 46475 g 61039 t 300 others  
ORIGIN

Query Match		10.9%;	Score 467.4;	DB 2;	Length 212601;
Best Local Similarity		78.8%;	Pred. No. 2.9e-90;		
Matches 738;		Conservative	0;	Mismatches 146;	Indels 53; Gaps 13;
QY	3341	ttttgattcaggtgaaagttaagccatattgtcttgatgatcagctgtgtgatgaatgt 3400			
Db	180590	TCTCCAAATCAGTGGAGGTTAAGCCATATGCTTGGATGACCGAGCTGTGTATCAATGT 180649			
QY	3401	cagggggccggtgtgggggaaattgtccatttttctgtgtgtaattacctgtctg 3460			
Db	180650	CAAGGGCCCGTGTGGGGGAAATTTGCTCCATTTTCTGCTAAATGTTACCGTCTG 180709			
QY	3461	cagtattactgtgaattgtctgggctgtctatccattctcgtctggtggcagggaattccac 3520			
Db	180710	CAGTATTACTGTGAATTTGCTGGGCTGCTATTCACTCTCGTCTGGCAGAGAATTCAC 180769			
QY	3521	aagccctggtgaaggaagcggtgaccgccctcgccatatttccctgcgtggaact-- 3578			
Db	180770	AAGCCCTGCTGAAGGAAGGTGCTGACCGCCCTCGGCATATTTCAATTCGCTGGAAC TGA 180829			
QY	3579	--aaagataaactgaagtctcattttcaaggccctcagaataagtcactcttcttcat 3636			
Db	180830	AGAATGAATGACTCCAGAGCTCATTTGCAGGCCTCAGAGAAGTGCCTCTTTCTTAAT 180889			
QY	3637	cttgacccttctccaaactcttcacgtgcatgtcccttttgttagcagctgtgaactta 3696			
Db	180890	TC-CAACTCTTCTCAGCTCTTCACAGTGGCATGTCC-TGCACAGCAGTGTGAAC-TA 180946			
QY	3697	actatagtaataaagaaagacatataataggtgtttgttagattcttctgtgcac 3756			
Db	180947	TCAGTGTGAGATGAAGAAAGAACCACTACACATAGGTATTTGTAGACTCTCGTGCAC 181006			
QY	3757	tgcaacaatatgaactcct--tttctgattgccatcggttgcatggaagttt-att 3813			
Db	181007	TGAACAACAGTATGTTCAACTCCATTTCACATTGCCATTATGTTGCATGCAAGTTTTAAT 181066			
QY	3814	ctctgtttctggaacccagagatccaaactccttcgcacatttcttcttagaggaga 3873			
Db	181067	CTCTGTTTTCGCGAGGCCAGAGAGATCTAAACTGCTGTAGCATTTCTTAGAG- --- 181122			
QY	3874	gagagaataattaaaagagaatgaacaatatagattttgggtttttaataataattat 3933			
Db	181123	GAGAGAAATATTAAACAGAAATGAGACAATA- -----TTTGTAGGTTTTTAATTGC 181172			
QY	3934	tgttaataataacatatagaatacttttattaaataacacatgcgaacaataacacta 3993			
Db	181173	TGTAAATGATATAACATATAAGAAAACCTTTATTGAGATAATC--ACCAACAGTAACACTA 181230			
QY	3994	tcgggtctatctgacagtttttccccagggaaagtgtcttttgccttttcttcttttt 4053			
Db	181231	TCATCTGTCTGACAGCTCTCCCCGAGAAAGTGTCTTGCCAGTCTTTCTT- ---- 181285			
QY	4054	tttttttccatctttttgtctctctcttttttccatcccttttttaattttttaacag 4113			
Db	181286	-----TCATTCTCTCTCTCTCTTTTCTTCTATCCCTTTTA--TTCTTTAACAG 181331			
QY	4114	caatgagaagttaacaatttttaa-tggaaagagcatgttagagcaacaacaaatgcata 4172			
Db	181332	CAATGGAGGAAGTTAACAGTATTATAACAGAAAGAGTGTGTGAGCAAGCAAAATGCATG 181391			
QY	4173	agcaagactgagcagcattataattttcagggttttgagcgtgaacataatttcat 4232			
Db	181392	AGCAAGATTGAGCAG-----TGTAAACCATAGAGCTTTGAGGCTGTATGATAGCTTGAT 181444			
QY	4233	tatccctcaaaaagttaccaccacacatcagaaaaaaa 4269			
Db	181445	CATGCTCAAAATGCCCTTCATACACAGAAAAATAAA 181481			
RESULT	7				
AL669946/c		249657 bp	DNA	linear	HTG 30-JAN-2002
LOCUS	AL669946				

DEFINITION	Mus musculus chromosome 11 clone RP23-71G18, *** SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION	AL669946
VERSION	AL669946.2 GI:18477046
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (sites)
TITLE	Burton, J.
JOURNAL	Direct Submission
COMMENT	Submitted (18-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerv@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Feb 1, 2002 this sequence version replaced gi:18250611. ----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: humquerv@sanger.ac.uk ----- Project Information Center project name: BM71G18 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Chemistry: Dye-terminator Big Dye; 100% of reads Consensus quality: 245751 bases at least Q40 Consensus quality: 246574 bases at least Q30 Consensus quality: 247123 bases at least Q20 Insert size: 247957; sum-of-contigs Quality coverage: 236934; 4.0% error; agarose-fp Quality coverage: 9.57x in Q20 bases; sum-of-contigs Quality coverage: 10.17x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.
FEATURES	Location/Qualifiers
source	1. 249657 /organism="Mus musculus" /db_xref="taxon:10090" /chromosome="11" /clone="RP23-71G18" /clone_lib="rPCI-23"
misc_feature	1. 12488 /note="assembly_fragment:01452 fragment_chain:1 clone_end:SP6 vector_side:left" 12589..18640 /note="assembly_fragment:00229 fragment_chain:1" 18741..20809 /note="assembly_fragment:03184 fragment_chain:1" 20910..24121 /note="assembly_fragment:04869 fragment_chain:1" 24222..32488 /note="assembly_fragment:04953 fragment_chain:1" 32589..51507 /note="assembly_fragment:05426 fragment_chain:1" 51608..64155 /note="assembly_fragment:03730 fragment_chain:1" 64256..72075 /note="assembly_fragment:01963 fragment_chain:1" 72176..79866 /note="assembly_fragment:05385

```
misc_feature      fragment_chain:1"
79967..85184
/note="assembly_fragment:00077
fragment_chain:1"
85285..103614
/note="assembly_fragment:00749
fragment_chain:1"
103715..132151
/note="assembly_fragment:03729
fragment_chain:1"
132252..147168
/note="assembly_fragment:03424
fragment_chain:1"
147269..152571
/note="assembly_fragment:04053
fragment_chain:1"
152672..177942
/note="assembly_fragment:04381
fragment_chain:1"
178043..190101
/note="assembly_fragment:00306
fragment_chain:1"
190202..217559
/note="assembly_fragment:04881
fragment_chain:2"
217660..249657
/note="assembly_fragment:03901
fragment_chain:2"
BASE COUNT      70686 a 54329 c 54183 g 68743 t 1716 others
ORIGIN
Query Match      10.9%; Score 467.4; DB 2; Length 249657;
Best Local Similarity 78.8%; Pred. No. 2.9e-90;
Matches 738; Conservative 0; Mismatches 146; Indels 53; Gaps 13;
QY 3341 ttttgatgcagggtgaagtttaagccatctgtcttggatgcagctgtgtgatgaatgt 3400
Db 210842 TCTCCAATGCAGTGGAGGTAAAGCCATATGTTGGATGACCAGCTGTGTGAATGT 210783
QY 3401 cagggggccgtgtgggggaaattgctccatttctctgtctgaatgttaacctgtcg 3460
Db 210782 CAAGGGCCCGTGTGGGGGAAATTTGCTCATTTTCTGTCTGAATGTACCTGTCTG 210723
QY 3461 cagtattactgtgaattgtgggctgtatccattctgtctgctggcagggaattccac 3520
Db 210722 CAGTATTACTGTGAATATTGCTGGGCTGCTATTCTACTCTCGTGGCAGAGATTCCAC 210663
QY 3521 aagcccttggtgaaggaaggcggtgaccgccctcgccatattcattccgtggaact-- 3578
Db 210662 AAGCCCTGGTGAAGGAAGGTGGTGACCGCCTCGGCATATTTCATTCGCTGGAAC 210603
QY 3579 --aaggataactgcagtgctcatttgcagccctcagaataagtcgactcttctgttcat 3636
Db 210602 AGAATGAATGACTCCAGAGCTCATTTGCGAGGCTCAGAAGAAGTGCCTCTTTGTTAAT 210543
QY 3637 tctgacctctctcaacctctcaactgtgcatgtctcttctgtagcagctgtgaactta 3696
Db 210542 TC-CAACTCTTCTCAGCCTCTTCACACTGGCATGTCC-TGCACAGCAGTGTGTAACTA 210486
QY 3697 actatagtaataatgaagaatgaacctataatagtggttttttagattcttctgtcacc 3756
Db 210485 TCAGTGGTGAATGAAGAAAGAACACCTACACATAGGTATTTTGTAGACTCTCGTGTAC 210426
QY 3757 tgaacaacaatatgaactcct--ttttcgattgcatcgggttgcattggaagttttt-att 3813
Db 210425 TGCAAACAGTAGTGTTCAACTCCATTTTCATATTGCCATATGTTGCATGCAGGTTTAAT 210366
QY 3814 ctctgttttctggaaccagagaggtaccacactctcctgcacatttctcttagaggaga 3873
Db 210365 CACTGTGTTTCTGCGGCCGAAGAGGATCAAACTGCTGTAGCATTTTCTTAAG---- 210308
QY 3874 gagagaaatattaaagagaataatgaacaatatagagtattttgggtttttaataaattat 3933
```

```
Db 210309 GAGAGAAATATTAAACAGAGAAATGAGACAATA-----TTTTGAGGTTTAAATTGC 210260
QY 3934 tgttaataataacacatataagaataactttttattataaataaacatgcacaacaataacacta 3993
Db 210259 TGGTAATGATATAACATATAAGAAAACTTTATTGAGATAATC--ACCAACAGTAACACTA 210202
QY 3994 tcggtctatctgcacagtttttcccccagggaagtgttttgccttttctcttctttttttt 4053
Db 210201 TCATCTGTTCTGACACAGTCTCCCCCAGGAAAGTGTCTGTGCCAGTCTCTTCTT 210145
QY 4054 ttttttctcatotttttgtttctctctctttttccatcccttttttaatttttttaacag 4113
Db 210146 -----TCATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 210101
QY 4114 caatggagggaagttaacaatttttaa-tggaaagacagctgttagagcaacaacaataacata 4172
Db 210100 CAATGGAGGAAGTTAAACAGTATTAAACAGGAAAGATGTATGCAGACGACGAATAATG 210041
QY 4173 agcaagactgcagcagcattataataatttttcaggggttttgcagctgaacataattcat 4232
Db 210040 AGCAAGATTGAGCAG-----TGTAACCATAGAGCTTTGAGGCTGTATCATAGCTTGT 209988
QY 4233 tatccctcaaaaagttaccaccacatcagaaaaaaa 4269
Db 209987 CATGCCTCAAAATGCCTTCATACACACAGAAAAATAAAA 209951
RESULT 8
AB023157
LOCUS      Homo sapiens mRNA for KIAA0940 protein, complete cds.
DEFINITION
AB023157
ACCESSION
AB023157.1 GI:4589523
VERSION
KEYWORDS
SOURCE      Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
SK plus clone:hh04894.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirose,M.,
Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
Prediction of the coding sequences of unidentified human genes.
XIII. The complete sequences of 100 new cDNA clones from brain
which code for large proteins in vitro
DNA Res. 6 (1), 63-70 (1999)
99246063
MEDLINE
2 (bases 1 to 5715)
Ohara,O., Nagase,T. and Kikuno,R.
Direct Submission
Submitted (04-FEB-1999) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
292-0812, Japan (E-mail:cdhainfo@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
FEATURES
Location/Qualifiers
1..5715
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hh04894"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
/dev_stage="adult"
90..2144
/gene="KIAA0940"
90..2144
/gene="KIAA0940"
/codon_start=1
/product="KIAA0940 protein"
/protein_id="BAA76784.1"
/db_xref="GI:4589524"
/translation="MQDILLMDKSKTPQPPQOORQOQOPESSVSEAPSTPLSSET"
gene
CDS
```

```
PKPEENSAVPAISPAAPAPNGPDKMOMESPLILGLSFHPQPPQPPPPQPPAAPCAS
LSPFSFGTSTGTTNAVEDSFQGLTPVNGTMLFONFPHVNPVFGGTFSPQIGLAQT
QHQQQPPAPAPQPAQPAQPAQPPQRRPASPQAPYACRAAAAAYGHQPIAMTS
KSSSAVAAAAAASASSSWNTHQVNAASAPSNPWGGLQAGPRRAGVGV
GVGVGVSPLNPLISPLKFPFSNNVIAPPKFPPRAAPLTSKSMEDNAFTDNGSNLLPF
ODRSYDFNLHLSNLSLMDIRTDHEPLKGRMFINFHPGTDITMALNSRSLPPE
EDAFLODSDHQAALSSGLSPYRCQNGVERVSRKVFVGGPLPDIDEIDETASPRRF
GLIVVDWPHKAESKSYFPPKGVAFLLFOEVSVOALIDACLEEDKMLKLYCSPIIKD
KPQVIRPNWLSUSDFVMDGSQPLDPKRTIFVGVPRPLRAVELAMIMKLYGVCTAG
IDIDPELTKPAGRAVAFSNQSYIAAISAREVLOHNDIKRVEKPKPYVLDQMCDE
CQGTGKGKFAPEFCANVTCLQOYCEYCWASIHRSRAGREFHPLVKEGGDRPHVFFR
WS"
BASE COUNT      1571 a  1201 c  1211 g  1732 t
ORIGIN

Query Match      10.1%; Score 430; DB 9; Length 5715;
Best Local Similarity 77.0%; Pred. No. 2.7e-82;
Matches 539; Conservative 0; Mismatches 155; Indels 6; Gaps 1;

Qy 33 gtcagttcactgtttccgatgaagaatgattcttgatgatgcccgtggagatcagc 92
Db 1243 GTCGATCTTCTCTTTCCCTTTGAAGATGCCCTCTCGGATGATAGCCATGGTGAACAAG 1302

Qy 93 ctcttcatagtggcgtgggttcacctcaactgcttcagtcaccagaaatgggagagagatgg 152
Db 1303 CCTTGTCTATCGGCTTAAGTTCCTCCACTCGCTGT-----CAAAATGGGAACGAGTAG 1356

Qy 153 aacgatattctcgaaggtttttagtggcgtgattgctccagacattgatgaagatgaga 212
Db 1357 AACGCTACTCTAGAAGGTGTTGTTGGAGGACTTCTCTCGATATATGATGAAGATGAGA 1416

Qy 213 tcacagtagtttctgcgtgttggtgcccctctgattgattgattgctgctcctaaagctgaga 272
Db 1417 TCACGTGCAGCTTTCGCGAGTGTGGAGCTCTCTGCTAGACTGAGCTGCTCAAAAGCTGAAA 1476

Qy 273 gcaaatcctattttccctcctaaaggtctatgctctcctgctgtttcagatgaagctctg 332
Db 1477 GCAAGTCTATTTCCTTCCTTAAGGCTATGCTCTTCTGCTGTTCCAGAGGAAGCTCAG 1536

Qy 333 tgcaggtctcattgctgcattggaagattgaagaagatggaataactctacccttgtgtatcaa 392
Db 1537 TACAAGCTTTGATAGATGCGCTGCGCTAGAGAAGATGGGAACCTACCTGTGTGTCTCAA 1596

Qy 393 gtccactatcaaggataaagccagtcagattcgccctggaatctcagtgacagtgact 452
Db 1597 GCCCCACATCAGGACAAACAGTGCGAATTCGACATGGAACCTAAGTGACAGTGACT 1656

Qy 453 ttgtgattggtgttcacagccacttgaccacgaaatactatttgggtgtgtgttc 512
Db 1657 TTGTAATGGATGTTCTTCAGCCTTTGGACCCACAGAAAACATCTTTGTGGGGAGTTC 1716

Qy 513 ctgacattacagagctgtgagactgctgagtgatgtaatgagtcgctatacggaggtgctg 572
Db 1717 CACGACCCCTTCGAGCTGTGTAATCGCAATGATGATGACCGCTTTGTACGGTGTGTCT 1776

Qy 573 gctacgtggtgattgataccagccctgagctaaaaatacccaaaagagctggagagattg 632
Db 1777 GCTATGCTGCATGTATACGACCCAGAGCTGAAGTACCCCAAGGTGCTGGCCGCGTGG 1836

Qy 633 cgttcttaatacacagaggttacatagctgtctatcagtcgcccttctgttcagctcagc 692
Db 1837 CATTCCTCAATCAGCAGAGTTACATTCGACCCATCAGCGCTCGCTTTGTGCAGCTTCAGC 1896

Qy 693 atggagagatagataaacagggttaagccttatactacattt 732
Db 1897 ACAATGACATTGCAAAACGGGTTTGAAGTAAGGCAATATGT 1936

RESULT 9
AC100917 AC100917 71171 bp DNA linear HTG 23-NOV-2001
LOCUS Mus musculus clone RP23-71G18, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
```

AC100917  
AC100917.1 GI:17059691  
HTG: HTGS\_PHASE0.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 71171)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Mus musculus, clone RP23-71G18  
Unpublished  
2 (bases 1 to 71171)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukigalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Hearford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,  
Jones,C., Kanat,A., Karatas,A., Kellis,C., LaRocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,  
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,N.

TITLE  
JOURNAL  
COMMENT

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L14574  
Center clone name: 71\_G\_18

\* NOTE: This record contains 87 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 718: contig of 718 bp in length  
719 818: gap of 100 bp  
819 1541: contig of 723 bp in length  
1542 1641: gap of 100 bp  
1642 2370: contig of 729 bp in length  
2371 2470: gap of 100 bp  
2471 3209: contig of 739 bp in length  
3210 3309: gap of 100 bp  
3310 4026: contig of 717 bp in length  
4027 4126: gap of 100 bp  
4127 4836: contig of 710 bp in length  
4837 4936: gap of 100 bp  
4937 5664: contig of 728 bp in length













gene

```
<10319..>29827
/feature="BcDNA:LD29930"
/feature="CG4999"
/feature="66E3-66E4"
/feature="FBan0004999"
/db_xref="FLYBASE:FBgn0029505"
join(23890..23955,24459..24557,25779..25907,28143..28502,
28563..28712)
/feature="BcDNA:LD29930"
/feature="BcDNA:LD29930 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0004999"
/db_xref="FLYBASE:FBgn0029505"
/protein_id="AAF50354.1"
/db_xref="GI:7295026"
/translation="MVFDGCVWCAKYLICIFNFVPLGTIIFFVGLWLVADKHSLLIA
LWLSERIEQTPQAELEQALVLLVIGVAGFFMSFLGILGAMRESKRLISTGYTF
LILLLEAVAGGAFKDKVRAESKNFLOTITTSYSGENVDATSLMNLGMLNFG
CCGINIDYHDDASPAWNGKNGRTIPDACCILDKVAKLVPRDEDCITNPDSNSFYKK
GCYEFTWELIRORELVIIVAIAGVIVHLVLIILAFALCKAFKYNDRU"
complement(join(<30826..31252,31329..31903,32168..32359,
32429..32701,32759..32966,33028..34517,34833..>34850))
/feature="CG5741"
/product="CT18050"
/db_xref="FLYBASE:FBan0005741"
/db_xref="FLYBASE:FBgn0035937"
complement(<30826..>34850)
/feature="CG5741"
/feature="66E3-66E4"
/db_xref="FLYBASE:FBan0005741"
/db_xref="FLYBASE:FBgn0035937"
complement(join(30826..31252,31329..31903,32168..32359,
32429..32701,32759..32966,33028..34517,34833..>34850))
/feature="CG5741"
/feature="CG5741 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0005741"
/db_xref="FLYBASE:FBgn0035937"
/protein_id="AAF50353.1"
/db_xref="GI:7295025"
/translation="MOVVKQVYHRRPAVQVSPHCRRPVQVQVEEAPWAGVAAAGDSN
SSGTVRLIKGVKTQITIEIKSGSLPPTAAPIIVADPPSTGWSNDPDSDTLCRLVNT
SISEPLDLDEHONGSDSVEAELLQHSISLGSQDVPVTAATGHGSRSLRGLAHI
RSVKAKQAAPAMNKDVSQSGSTVPATSVRTGFLRRNVVEAGSEPTSQOAEAI
ASGVKKRGILLARKDVEIESGVEVLEDMIPSTAVIHDSKDVANKRAEVDAGADA
DSRASLTPVAPGFTSLAISYKEGSDGRTSRISFFSNLSMVSQWREKHQGATA
FALPVLVLLIILLMLPVPDLRGVFATLLFTVMDCCGHLRALLEDLMFLTHPER
VAFAPNPKMPICEIPAVEBEHTIKTYAGWMEINSYDPNFTFSFLTRAVVYRLDGS
ILKMSGTNARIPKRWMEPPIDRHKIPFDHRSYDLRDCRIELPLGLANKRFFNRK
YPIQLIISNRDVPDQSVESQSEATVKTDAKEAKASSGTTTPGGEELVDPAG
TVNQDLQOLRNTVNPDKELNDITVPCGDEVRLLLFSRCDREKEDWYRRTNASRGLV
HEQDLQPMFARVEDSLQAAARQALLVITGMKGAGVKSSEDEISAGKNDETDQTP
DGQFETIEEDLDTPKESFDGMIANSVARNSPDYRFMAVYRRARKEDELEWK
GIDQSLFLPSGSGIYVWVLLGRCLYSWLHADMHEKLEKLVQKLNSIKLPSFMEEV
IITNIVGLSPGLCHRVSPMLDERGVADIVTYEGLAHITVTKLNLRLRSKTKS
PMATDNVPTSMLEPOSODLRSISEVNSENVLYDSDAESGSGSTSESPPCNPAPEN
AACTEFQNSPGNARLFIKIVDRIAAASNLFOYATELPYVQRAMENNNANITLRDLKG
LVARGILNPPSPSRVWSFGPRPLWISITKRPQVGDKSVDSIVTNVJESLCEAVN
KFLIYPMNVDFSIFLUSNPNFDDQNAAN"
complement(join(39649..42284,45592..45653,46647..47044,
47421..47523,50583..50980))
/feature="CG5735"
/product="CT17928"
/db_xref="FLYBASE:FBan0005735"
/db_xref="FLYBASE:FBgn0035938"
complement(join(39649..42284,45592..45653,46647..47044,
47421..47523,51697..52020))
/feature="CG5735"
/product="CT18021"
/db_xref="FLYBASE:FBan0005735"
/db_xref="FLYBASE:FBgn0035938"
complement(join(39649..42284,45582..42693))
/feature="CG5735"
```

mRNA

mRNA

mRNA

```
/product="CT17898"
/db_xref="FLYBASE:FBan0005735"
Query Match 7.1%; Score 304.6; DB 3; Length 287018;
Best Local Similarity 70.6%; Pred. No. 2.6e-55;
Matches 406; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
QY 140 ggggagagagtgaaacgatatcttcgaaagggtgttgtagcgagattgctccacacatt 199
DB 41455 GCGGATGGCAGCGCTCGCTTCGCCGGAAGGATATTCGTTGGCGGCTGCCACCGACATC 41396
QY 200 gatgaagatgagatcacacagctagtcttcgctctttggccctctgatgtgtagtgccct 259
DB 41395 GATGAGGATGAGATTACCACCTTCGTTTCGGCGCTTCGGGCCATTGTCGTCGATTGGCCA 41336
QY 260 cataaagctgagagcaaatccctattctctctctctctctctctctctctctctctctctct 319
DB 41335 CACAAGCGGAAATCCAAAGTCGATATTTCCGCCCAAGGGATATGCCCTCCTGCTGTCCAG 41276
QY 320 gatgaagctctgtgcagctctctcattgatgatgatgatgaagaagatgaagaactctac 379
DB 41275 GACGAGACGATGTGCAGCAGCTAATTGACTCGTCGATCACCAGATGAGGACAAGCTGAT 41216
QY 380 ctttgttatcaagtcctccactcaaggataaaggccagtcacagctcgcccttggaatctc 439
DB 41215 CTATGCGTTTCTTCGCGCAGCATCAAGGATAAGGCGAGTCAGATTCGTCCTTTGGCGCTG 41156
QY 440 agtgacagtgaactttgtgatgatggttcacagccacttgacccacgaataactatatt 499
DB 41155 GCGGATCGGAGCTATGTGTTGATGTACCATCTCAGTGCAGCCACGCAAAACGGTGT 41096
QY 500 gttgggtgttctcgcacattacagctgtggagcttgcatgtaagtgcggtc 559
DB 41095 GTGGCGCGGTGCGCCTCTCTGAAGCCTTCGAAGCTTGCAGTGCAGTATGATGATGATG 41036
QY 560 tacggaggtgtgtacgctgggtgatgataccgacccctgagctataaaataccacaaagga 619
DB 41035 TAGGGTGGAGTATGCTATGCTGGAATTGACACCGATCCGGAATTAAGTATCCAAAGGC 40976
QY 620 gctgggagagtgctgtctctctctctctctctctctctctctctctctctctctctctctct 679
DB 40975 GTGGAGCTGTGGCTTCTCGAATCAGCAGAGTACATAGCGGCCCTCTCAGCCAGATT 40916
QY 680 gttcactcagcatgagagatagataaacgggt 714
DB 40915 GTGCAGCTGCAQTGGCGATATAGACAACGGGT 40881
RESULT 14
AC009968/c 205307 bp DNA linear HTG 26-JUL-2000
LOCUS Homo sapiens chromosome 19 clone RP11-402N14, WORKING DRAFT
DEFINITION SEQUENCE, 22 unordered pieces.
ACCESSION AC009968
VERSION AC009968.6 GI:9454634
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
REFERENCE 2 (bases 1 to 205307)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 26, 2000 this sequence version replaced gi:8748919.
----- Genome Center -----
```





```
QY 1888 tcactcaacactcctcccggttcaagtgattcctccttagcctcccgagtagg 1947
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122959 TCACTGCAACTCTGCTCCAGGTTCAAGCAATTCCTGCTCAGTCACCTGAGTAC 122900
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1948 tgaattacaggcatataccaccatcccaacaatgtttgtatttttttagtgagacagg 2007
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122899 TGGGAATACAGCGCCACCACCAATCCCGAGCAATTTTGTGATTTTATAGTAGAGACAGG 122840
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2008 gtttcaccgtgtgttggtcagcgagtttcaaaactcctgcactcaaggagatccacctgcctc 2067
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122839 GTTTCACCATGTTGCCAGGCTGTTTTCGAACCTCCTGAGCTCAGGTGATCGGCCACCTTC 122780
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2068 acccctcaaatgtcgtggtattacagcatgacccaccaccatgcctgacctgtgt 2127
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122779 GGCTCCCAAGTGTGCGGATTAGATGTGAGCCACCTCGCTGCCCCAAAAGCTCTAA 122720
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2128 tttttgcacactaaaaataacctcatctcactcctccttattccaaataagtttcaga 2187
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122719 TTTTATG-----AGAAACTCTGAGGACAGAACTTGTAGTCAATTGTTAATGAATNN 122669
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2188 gctgtgggattggtcattagaaattcagactgaattgttctcctcgaatgaatactc 2247
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122668 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 122609
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2248 tgcaccagtgtcactcctgtagacattatgagcagcctagagccagcagaagccca 2307
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122608 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 122554
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2308 gtctctccttgcctgtctctcctcctggttcgtgcactctctctctctctcttctt 2367
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122553 TTCCTTCTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 122494
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2368 ttatttttttagtaaaaaatttttttagaggagggtcactcactcgtcaccagcgt 2427
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122493 TTTCTCTCTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 122434
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2428 ggaacacagaatcacatcatcactcactcactgtctctctctctctctctctctc 2483
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122433 GGACTCGATGGTCAATCTCGGCTCACTCAAGCTCCACCTCCAGGTTTCAGGCC 122378
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 15
AC004973/c
LOCUS AC004973 97037 bp DNA linear PRI 21-DEC-1999
DEFINITION Homo sapiens PAC clone RP5-113911 from Xq23, complete sequence.
ACCESSION AC004973
VERSION AC004973.1 GI:3694660
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 97037)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 97037)
AUTHORS Courtney,L., Langston,Y. and Drone,K.
TITLE The sequence of Homo sapiens PAC clone RP5-113911
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 97037)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 97037)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 97037)
```

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Waterston,R.  
Direct Submission  
Submitted (18-MAR-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
6 (bases 1 to 97037)  
Waterston,R.  
Direct Submission  
Submitted (21-DEC-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Oct 3, 1998 this sequence version replaced gi:3213024.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: sapiens@watson.wustl.edu  
----- Summary Statistics  
-----  
Center project name: H\_DJ1139101  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:

This sequence was generated from part of bacterial clone contigs of  
human chromosome X, constructed by the chromosome X mapping group  
at the Sanger Centre, Wellcome Trust Genome Campus, Hinxton, UK.  
Further information can be found at  
http://www.sanger.ac.uk/HGP/ChrX/

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-5, prepared by  
Pietter de Jong and coworkers at the Roswell Park Cancer Institute  
(http://bacpac.med.buffalo.edu) using the method described by  
Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from  
one male donor.  
The clone may be obtained either from Genome Systems, Inc.  
(http://www.genomesystems.com) or Research Genetics, Inc.  
(http://www.resgen.com); or from Pieter de Jong.  
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-555N2; the clone sequenced  
to the right is RP3-404F18, 200 bp overlap. Actual start of this  
clone is at base position 1 of RP5-113911; actual end is at 13269  
of RP3-404F18.

FEATURES

Location/Qualifiers  
1..97037  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="Xq23"  
/clone="RP5-113911"  
/clone\_lib="RPCI-5"  
530..801  
/rpt\_family="L2"  
830..1035  
/rpt\_family="Alu"  
2179..2349  
/rpt\_family="MER1\_type?"  
3434..3489  
/rpt\_family="MIR"  
3526..3739  
/rpt\_family="MIR"  
4681..4852  
repeat\_region  
repeat\_region  
repeat\_region  
repeat\_region  
repeat\_region  
repeat\_region

[illegible][illegible]

Search completed: September 1, 2002, 09:45:29  
Job time: 14874 sec

---





XX WPI; 2000-097328/08.  
DR P-ESDB; AAW90737.

XX DNA sequences preferentially expressed in IgA nephropathy patients,  
PT proteins encoded by them, and antibodies to those proteins -

XX Claim 1; Page 76-83; 180pp; Japanese.

XX This invention describes novel DNA sequences preferentially expressed in  
CC IgA nephropathy patients, and DNA sequences stringently hybridizing to  
CC them. Independent claims cover diagnostic reagents for IgA nephropathy  
CC incorporating the antisense sequences; the treatment of IgA nephropathy  
CC using the antisense sequences for mRNA inhibition; proteins associated  
CC with IgA nephropathy, containing sequences encoded by the DNA sequences;  
CC antibodies recognizing these proteins; the production of the proteins  
CC by culture of host cells transformed with DNA encoding them; diagnostic  
CC reagents for IgA nephropathy containing the antibodies; and compositions  
CC for the treatment of IgA nephropathy which contain the antibodies. The  
CC products of the invention can be used for the diagnosis and treatment of  
CC IgA nephropathy. This sequence encodes the human IgA  
CC nephropathy-associated protein INP303A which is described in the method  
CC of the invention.

XX SQ Sequence 4276 BP; 1149 A; 854 C; 854 G; 1419 T; 0 other;

Query Match 100.0%; Score 4276; DB 21; Length 4276;  
Best Local Similarity 100.0%; Pred. NO. 0;  
Matches 4276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ttctacggtttttccctggtttctattccaggctcaggttcaggtttccctggtggaaga	60
DB	1	ttctacggtttttccctggtttctattccaggctcaggtttccctggtggaaga	60
QY	61	tggattcttgatgagcggcggtgggagatcagctcttcacagtgccctgggttcactca	120
DB	61	tggattcttgatgagcggcggtgggagatcagctcttcacagtgccctgggttcactca	120
QY	121	ctgcttcagtcaccagaaatggggagagatggaaacagatattctcgaagaggtttgttag	180
DB	121	ctgcttcagtcaccagaaatggggagagatggaaacagatattctcgaagaggtttgttag	180
QY	181	cggattgctccagacattgatgaagatgagatcacagctagtttctgctgttgccccc	240
DB	181	cggattgctccagacattgatgaagatgagatcacagctagtttctgctgttgccccc	240
QY	241	tctgattgtggtggtcctcataaagctgagagcaaatcctatttctcctctaaaggcta	300
DB	241	tctgattgtggtggtcctcataaagctgagagcaaatcctatttctcctctaaaggcta	300
QY	301	tgcattctctggtttcaagatgaagctctgtgcaggtctctcatigatgcatgattga	360
DB	301	tgcattctctggtttcaagatgaagctctgtgcaggtctctcatigatgcatgattga	360
QY	361	agaagatggaataactctacctttgtgtatcaagtcaccatcataagataagcagatcca	420
DB	361	agaagatggaataactctacctttgtgtatcaagtcaccatcataagataagcagatcca	420
QY	421	gattcggccttggaaatccagtgacagtgactttgtgagtgagtggttcacagccactga	480
DB	421	gattcggccttggaaatccagtgacagtgactttgtgagtgagtggttcacagccactga	480
QY	481	ccacgaaaaactattttgtgtgtgttctcgcagcattacagagctgtgagcttgc	540
DB	481	ccacgaaaaactattttgtgtgtgttctcgcagcattacagagctgtgagcttgc	540
QY	541	gatggttaagtatcggtctacagagtggtgtcgaagctgggattgataccagaccctga	600
DB	541	gatggttaagtatcggtctacagagtggtgtcgaagctgggattgataccagaccctga	600
QY	601	gctaaaaatcccaaaaggagctgggaggttgcgttctctataacacagagttacatagc	660
DB	601	gctaaaaatcccaaaaggagctgggaggttgcgttctctataacacagagttacatagc	660

Db	501	gctaaaaatcccaaaaggagctgggaggttgcgttctctataacacagagttacatagc	560
QY	661	tgctatcagtgcccgctttgttcagctgcagcatgagagatagataaaacgggttaagcct	720
Db	661	tgctatcagtgcccgctttgttcagctgcagcatgagagatagataaaacgggttaagcct	720
QY	721	tatactacattttggaaaaattctagaaatggtcctcctaaatggtgattaccaatattag	780
Db	721	tatactacattttggaaaaattctagaaatggtcctcctaaatggtgattaccaatattag	780
QY	781	aacgggagcattttatgacataaagtgcagctgacaattttgctctagagtttaatta	840
Db	781	aacgggagcattttatgacataaagtgcagctgacaattttgctctagagtttaatta	840
QY	841	tggctctataatacaatgaataatgctcctatgaattctcttcttcttcttcttcttctt	900
Db	841	tggctctataatacaatgaataatgctcctatgaattctcttcttcttcttcttcttctt	900
QY	901	agcctaatacgaac	960
Db	901	agcctaatacgaac	960
QY	961	atctcaatccatccgctccattccattccattccattccattccattccattccattcc	1020
Db	961	atctcaatccatccgctccattccattccattccattccattccattccattccattcc	1020
QY	1021	gaatccctttgtctgtttatatacaagatatacaaacaggaatgccttgcgttccctgc	1080
Db	1021	gaatccctttgtctgtttatatacaagatatacaaacaggaatgccttgcgttccctgc	1080
QY	1081	cctttttttgtttgttttttttaactcctgggacataagggaagacctcagaagccctatt	1140
Db	1081	cctttttttgtttgttttttttaactcctgggacataagggaagacctcagaagccctatt	1140
QY	1141	ctcaatgaattgtactcacacagatttcttttttttttttttttttttttttttttttt	1200
Db	1141	ctcaatgaattgtactcacacagatttcttttttttttttttttttttttttttttttt	1200
QY	1201	acctctcacagatttattccttagcttgcgttgcgttgcgttgcgttgcgttgcgttgc	1260
Db	1201	acctctcacagatttattccttagcttgcgttgcgttgcgttgcgttgcgttgcgttgc	1260
QY	1261	ttagggcaagaagtctcctcatgagttatttctcctagcagatagaaactgtatcact	1320
Db	1261	ttagggcaagaagtctcctcatgagttatttctcctagcagatagaaactgtatcact	1320
QY	1321	tgcagtgactactcagagtggtgccttggaactgcagctccagctcgttaaaacttagtt	1380
Db	1321	tgcagtgactactcagagtggtgccttggaactgcagctccagctcgttaaaacttagtt	1380
QY	1381	agtgcagataggaatttagaccagaatgtaatacaaccacattactcgggacacattgtt	1440
Db	1381	agtgcagataggaatttagaccagaatgtaatacaaccacattactcgggacacattgtt	1440
QY	1441	gtccagctggcgatgttttttttccatgaagacctttatttcagtgaggggaagcaatatt	1500
Db	1441	gtccagctggcgatgttttttttccatgaagacctttatttcagtgaggggaagcaatatt	1500
QY	1501	atttatttttggggtcaccttttttctcagtgacacactggcacttttcagtcagctga	1560
Db	1501	atttatttttggggtcaccttttttctcagtgacacactggcacttttcagtcagctga	1560
QY	1561	ctttgatataccatcactcctgaggaattgtctaaaaatagattgttttatcgttgttctc	1620
Db	1561	ctttgatataccatcactcctgaggaattgtctaaaaatagattgttttatcgttgttctc	1620
QY	1621	tcaattcaaatgtataaaatcatcaagctagtagcagtttttgcctttttcatgttcatgt	1680
Db	1621	tcaattcaaatgtataaaatcatcaagctagtagcagtttttgcctttttcatgttcatgt	1680
QY	1681	catgtacagttctactcactcctggcagtaaaaaataaataagatagtggtgcattcctaca	1740
Db	1681	catgtacagttctactcactcctggcagtaaaaaataaataagatagtggtgcattcctaca	1740

QY 1741 acttgtaattattaaagagaaaaagtatctgttctatttctaagcatgggggaggaacaag 1800  
DB 1741 acttgtaattattaaagagaaaaagtatctgttctatttctaagcatgggggaggaacaag 1800  
QY 1801 attagtatgttaacatgcctactttgtttgtttgagatggagtcctctccgtccaccag 1860  
DB 1801 attagtatgttaacatgcctactttgtttgtttgagatggagtcctctccgtccaccag 1860  
QY 1861 gctggagtcagtggttacagttctcagctcactcccaacctctccctcccggttcaagtga 1920  
DB 1861 gctggagtcagtggttacagttctcagctcactcccaacctctccctcccggttcaagtga 1920  
QY 1921 ttctctctgcttagctcccgagtagtggaattacagagcatataccaccatgcccaca 1980  
DB 1921 ttctctctgcttagctcccgagtagtggaattacagagcatataccaccatgcccaca 1980  
QY 1981 aatgtttgtatttttagtgagacaggggtttcacogtgttgctcagggccagtttcaaaact 2040  
DB 1981 aatgtttgtatttttagtgagacaggggtttcacogtgttgctcagggccagtttcaaaact 2040  
QY 2041 cctgacctcaaggatccaacctgcctcacccctcaaaagtgctgggattacagggcatgag 2100  
DB 2041 cctgacctcaaggatccaacctgcctcacccctcaaaagtgctgggattacagggcatgag 2100  
QY 2101 ccaccacacatgcctggcctacttgggttttttatgcacactaaaaataacctacatctca 2160  
DB 2101 ccaccacacatgcctggcctacttgggttttttatgcacactaaaaataacctacatctca 2160  
QY 2161 ctgcctatttccaacataagtttcagagctgtgggattgggtcattagaaaattcagactga 2220  
DB 2161 ctgcctatttccaacataagtttcagagctgtgggattgggtcattagaaaattcagactga 2220  
QY 2221 atttggttccctctgaaatgaaatccttggcccagtggttcactgtcactctgtagacatta 2280  
DB 2221 atttggttccctctgaaatgaaatccttggcccagtggttcactgtcactctgtagacatta 2280  
QY 2281 tggagcagcctagagcccaagccagtgctctccttatgcctctccttctctctgggcttc 2340  
DB 2281 tggagcagcctagagcccaagccagtgctctccttatgcctctccttctctctgggcttc 2340  
QY 2341 gtagacactctctctctctctcttttgtaoitttttttttagtttaaaaaatttttttttag 2400  
DB 2341 gtagacactctctctctctctcttttgtaoitttttttttagtttaaaaaatttttttttag 2400  
QY 2401 ggagggctcaactctgtcaccagctggagcacagaatcaacaatcactgactcaactgat 2460  
DB 2401 ggagggctcaactctgtcaccagctggagcacagaatcaacaatcactgactcaactgat 2460  
QY 2461 gttctctcctctttgttcatggtctaactcttggtcaggattccttctcagagctgggtggc 2520  
DB 2461 gttctctcctctttgttcatggtctaactcttggtcaggattccttctcagagctgggtggc 2520  
QY 2521 accagtgctggtagacagcctgtgtaagggagtttcagoccatgaatctctccagactaaa 2580  
DB 2521 accagtgctggtagacagcctgtgtaagggagtttcagoccatgaatctctccagactaaa 2580  
QY 2581 aataaacagctcttttctagctgataattataaacagggtagctgttaatgcttgaag 2640  
DB 2581 aataaacagctcttttctagctgataattataaacagggtagctgttaatgcttgaag 2640  
QY 2641 gttcacatgacaggttggccgatgaacgcgtggaacagggccagtttttagaaattcacct 2700  
DB 2641 gttcacatgacaggttggccgatgaacgcgtggaacagggccagtttttagaaattcacct 2700  
QY 2701 ctgacttttagactcaggtgaaccattcttactgagaagaacaagcagggtttttagac 2760  
DB 2701 ctgacttttagactcaggtgaaccattcttactgagaagaacaagcagggtttttagac 2760  
QY 2761 tgtgaatcctatgctgatcttttttttttttttttaacagagttccaggttttgattat 2820  
DB 2761 tgtgaatcctatgctgatcttttttttttttttttttttttaacagagttccaggttttgattat 2820

QY 2821 aaccaacaatgttacactataaataagaaacacagccagcaggttttttaagacagctcag 2880  
DB 2821 aaccaacaatgttacactataaataagaaacacagccagcaggttttttaagacagctcag 2880  
QY 2881 aatcttggacgcagtagtcaggcatcttcacaccgacttgaatatgtgaagtcagttgt 2940  
DB 2881 aatcttggacgcagtagtcaggcatcttcacaccgacttgaatatgtgaagtcagttgt 2940  
QY 2941 gtgaaacttgatcatcttagttgatttgttttaaatatgatccacatatgacaaaaa 3000  
DB 2941 gtgaaacttgatcatcttagttgatttgttttaaatatgatccacatatgacaaaaa 3000  
QY 3001 tccagatcccaataataaatgagggtttatgtctatgaataatctcctgtgggttaat 3060  
DB 3001 tccagatcccaataataaatgagggtttctagtctatgaataatctcctgtgggttaat 3060  
QY 3061 ctcataacattcttagtctaaacagtttgcttcacttcactgatgtctgtctcaaatcccttt 3120  
DB 3061 ctcataacattcttagtctaaacagtttgcttcacttcactgatgtctgtctcaaatcccttt 3120  
QY 3121 tccittaaagagatgtttatttaataagaaaaaaatgtaaaatgatagataaaaaagcc 3180  
DB 3121 tccittaaagagatgtttatttaataagaaaaaaatgtaaaatgatagataaaaaagcc 3180  
QY 3181 ttactagttcttaaaagatgaactatccatttccagttaaatgaataattagtccctcc 3240  
DB 3181 ttactagttcttaaaagatgaactatccatttccagttaaatgaataattagtccctcc 3240  
QY 3241 tctttgggcaccttggacacagattcattcagatagtggttggaatgtacatgtatggtta 3300  
DB 3241 tctttgggcaccttggacacagattcattcagatagtggttggaatgtacatgtatggtta 3300  
QY 3301 agcattgctggcctagtcactgaaataatgtaaaactctatttttgattgaggtgggaat 3360  
DB 3301 agcattgctggcctagtcactgaaataatgtaaaactctatttttgattgaggtgggaat 3360  
QY 3361 taagccatatgtcttgatgatcagctgtgtgataatgacagggccgctgttggggg 3420  
DB 3361 taagccatatgtcttgatgatcagctgtgtgataatgacagggccgctgttggggg 3420  
QY 3421 gaaatttgctccattttctgtgctaaatgtttacctgtctgcagtattactgtgaatttg 3480  
DB 3421 gaaatttgctccattttctgtgctaaatgtttacctgtctgcagtattactgtgaatttg 3480  
QY 3481 ctggggctctatccaattctcgtgctggcagggaaattccacaagccctgtggaaggaag 3540  
DB 3481 ctggggctctatccaattctcgtgctggcagggaaattccacaagccctgtggaaggaag 3540  
QY 3541 cggtagccgcctcggcataatttccattccgtggaactaaagataaactcagtgctcat 3600  
DB 3541 cggtagccgcctcggcataatttccattccgtggaactaaagataaactcagtgctcat 3600  
QY 3601 tttcaggcctcagaataaagtgcaactctctgttctcattctgaccccttcccaacctctc 3660  
DB 3601 tttcaggcctcagaataaagtgcaactctctgttctcattctgaccccttcccaacctctc 3660  
QY 3661 acgctggcatgtcccttttttagcagctgtgaacttaactatagtaataatgaaagaatga 3720  
DB 3661 acgctggcatgtcccttttttagcagctgtgaacttaactatagtaataatgaaagaatga 3720  
QY 3721 cctataataataggtgttttttagattcttctgaccccttcccaacctctc 3780  
DB 3721 cctataataataggtgttttttagattcttctgaccccttcccaacctctc 3780  
QY 3781 cgtattgccatcggttgcaggaagttttattctctgttttctgtggaacccaagagga 3840  
DB 3781 cgtattgccatcggttgcaggaagttttattctctgttttctgtggaacccaagagga 3840  
QY 3841 tccaaaactcctcaacattttcttagagagagagagaataatataaagagaataagaa 3900  
DB 3841 tccaaaactcctcaacattttcttagagagagagagaataatataaagagaataagaa 3900  
QY 3901 caatagagtagttttgggttttttaattaaattattgtttaataataacataataagaatc 3960



```
|||||
Db 3901 caatagagcatcttgggttttaattaaattattgttaataataataacataagaataac 3960
QY 3961 ttttataaaaaaaccaatgcaacaataacactatcggtctctatctgacaggttttccccc 4020
Db 3961 ttttataaaaaaaccaatgcaacaataacactatcggtctctatctgacaggttttccccc 4020
QY 4021 gggaaagcgttttgccctttctctctctctctctctctctctctctctctctctctctct 4080
Db 4021 gggaaagcgttttgccctttctctctctctctctctctctctctctctctctctctctct 4080
QY 4081 ctttttccatcccttttaatttttttaacagcaatggaggaagttaacaatttttaatt 4140
Db 4081 ctttttccatcccttttaatttttttaacagcaatggaggaagttaacaatttttaatt 4140
QY 4141 ggaagagcatgttagagcaacaacaatgcataagcaagaactgagcagcattataattaat 4200
Db 4141 ggaagagcatgttagagcaacaacaatgcataagcaagaactgagcagcattataattaat 4200
QY 4201 tttcagggttttgaggcgtgaacataatttcattatccctcaaaaagtttacaccacatca 4260
Db 4201 tttcagggttttgaggcgtgaacataatttcattatccctcaaaaagtttacaccacatca 4260
QY 4261 gaaaaaaaaaaaaa 4276
Db 4261 gaaaaaaaaaaaaa 4276

RESULT 2
ABA62438
ID ABA62438 standard; DNA; 544 BP.
AC ABA62438;
XX
DT 01-FEB-2002 (first entry)
DE Human foetal liver single exon nucleic acid probe #10743.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
PD
PE 09-AUG-2001.
PF 30-JAN-2001; 2001WO-US000669.
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
XX Claim 1; SEQ ID NO 10743; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
```

```
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at fip.wipo.int/pub/published\_pct\_sequences.
XX
SQ Sequence 544 BP; 128 A; 113 C; 125 G; 178 T; 0 other;
```

```
Query Match 12.7%; Score 544; DB 22; Length 544;
Best Local Similarity 100.0%; Pred. No. 3e-91;
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 3301 agcattgctggcctagtcactgaaaaatgttaaactcttattttgttgagtggaagt 3360
Db 1 agcattgctggcctagtcactgaaaaatgttaaactcttattttgttgagtggaagt 60
QY 3361 taaccatattcttgatgatcagctgtgtgatgaatgcaggggcccgctgtggggg 3420
Db 61 taaccatattcttgatgatcagctgtgtgatgaatgcaggggcccgctgtggggg 120
QY 3421 gaaattgtccatttttctgtgctaattgttacctgtctgcagtattactgtgaattg 3480
Db 121 gaaattgtccatttttctgtgctaattgttacctgtctgcagtattactgtgaattg 180
QY 3481 ctgggtgctatccattctctgtgctgaggggaattccacaagcccctggtgaaggagg 3540
Db 181 ctgggtgctatccattctctgtgctgaggggaattccacaagcccctggtgaaggagg 240
QY 3541 cggtgaccgcctcgccatatttccattccctggaactaaaggataactgcagtgtccat 3600
Db 241 cggtgaccgcctcgccatatttccattccctggaactaaaggataactgcagtgtccat 300
QY 3601 tttcaggcctcagaataagtgcaactctctgttcaattctgaccccttcccaacctcttc 3660
Db 301 tttcaggcctcagaataagtgcaactctctgttcaattctgaccccttcccaacctcttc 360
QY 3661 acgctggcatgtcctttgtagcagctgtgaacttaacttagtataatgaaagaatga 3720
Db 361 acgctggcatgtcctttgtagcagctgtgaacttaacttagtataatgaaagaatga 420
QY 3721 cctataataagggtttttgttagattctgtgctacgtcaacaataatgaactcctttt 3780
Db 421 cctataataagggtttttgttagattctgtgctacgtcaacaataatgaactcctttt 480
QY 3781 cgtattgccatcggttgcatggaagttttattctctgtttgtctggaaccaagagga 3840
Db 481 cgtattgccatcggttgcatggaagttttattctctgtttgtctggaaccaagagga 540
QY 3841 tcca 3844
Db 541 tcca 544

RESULT 3
AAI42417
ID AAI42417 standard; DNA; 544 BP.
XX
XX AC AAI42417;
XX
XX 17-OCT-2001 (first entry)
XX
XX DE Probe #11103 used to measure gene expression in human placenta sample.
XX
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157272-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000663.
XX
```

PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488897/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human placenta -  
XX  
XX Claim 25; SEQ ID No 11103; 654pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes (SENP).  
CC The present sequence is one such probe. The probes are useful for  
CC producing a microarray for predicting, measuring and displaying gene  
CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders.  
XX  
SQ Sequence 544 BP; 128 A; 113 C; 125 G; 178 T; 0 other;

Query Match 12.7%; Score 544; DB 22; Length 544;  
Best Local Similarity 100.0%; Pred. No. 3e-91;  
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3301 agcattgctggcctagctacactgaaatgaaactcttattttgattgagtggaagt 3360  
Db 1 agcattgctggcctagctacactgaaatgaaactcttattttgattgagtggaagt 60  
QY 3361 taagccatatgttgatgacagctgtgtaataatcagggcccggttgaggag 3420  
Db 61 taagccatatgttgatgacagctgtgtaataatcagggcccggttgaggag 120  
QY 3421 gaaattgtccattttctgtcctaattgtaacctgtctgcagattactgtgaatttg 3480  
Db 121 gaaattgtccattttctgtcctaattgtaacctgtctgcagattactgtgaatttg 180  
QY 3481 ctgggctgtatccattctcgtctggcagggaattccacaagccctggtgaaggagg 3540  
Db 181 ctgggctgtatccattctcgtctggcagggaattccacaagccctggtgaaggagg 240  
QY 3541 cggtgaccgcctcgccatatttcattccgctggaactaaaggataactgcagtgctcat 3600  
Db 241 cggtgaccgcctcgccatatttcattccgctggaactaaaggataactgcagtgctcat 300  
QY 3601 ttccaggcctcagaataaagtcactcttctgttcattctgcagcccttctcaccctcttc 3660  
Db 301 ttccaggcctcagaataaagtcactcttctgttcattctgcagcccttctcaccctcttc 360  
QY 3661 acgtgcatgctctttgtgacagctgtgaaacttaactatagataaagaagaatga 3720  
Db 361 acgtgcatgctctttgtgacagctgtgaaacttaactatagataaagaagaatga 420  
QY 3721 cctataataggtgtttgttagattctgtgcactgcgaacaaatagactccctttt 3780  
Db 421 cctataataggtgtttgttagattctgtgcactgcgaacaaatagactccctttt 480  
QY 3781 cgtattgccatcgggttgcatggaagttttattctctgttctgttggaacccaagagga 3840  
Db 481 cgtattgccatcgggttgcatggaagttttattctctgttctgttggaacccaagagga 540  
QY 3841 tcca 3844  
Db 541 tcca 544

RESULT 4

AA573462  
ID AAS73462 standard; cDNA; 1290 BP.  
XX  
AC AAS73462;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
XX DNA encoding novel human diagnostic protein #9266.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX P-PSDB; ABG09275.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -  
XX  
XX Claim 1; SEQ ID No 9266; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS64197-AAS94564 represent novel human  
XX diagnostic coding sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1290 BP; 345 A; 273 C; 281 G; 391 T; 0 other;

Query Match 8.3%; Score 356.4; DB 23; Length 1290;  
Best Local Similarity 72.9%; Pred. No. 1.2e-56;  
Matches 529; Conservative 0; Mismatches 186; Indels 11; Gaps 5;  
QY 33 gtcagcttcctcactgtttccgattggaagatggattcttgatgagccggtgggatcagc 92  
Db 137 gtcggttcttcctcttccaaatagatgaggtctgttgatgaggtcacagtgtatca-- 194  
QY 93 ctcttcattagtgccctgggttcacctcactcactcagtcaccacagatggggagagtg 152

Db 195 -agttggaggttttaattcaaccaccctgttattcagctcaccacaaatggagagcgaatag 253  
QY 153 aacgatatctcogaaggtgttttagcggtattgcctccagacatgatgaagatgaga 212  
Db 254 aacgctctctcgaagaagttttgtgtgtgtcttccctccagatattgatgaagatgaaa 313  
QY 213 tcacagctagtttctgtctgttggccctctgattgtggttgccctcataaaagctgaga 272  
Db 314 taactgcctagcttcgaagaatttggcccttggtagtagattggcctcataaagcagaaa 373  
QY 273 gcaaatctatttctctcctaaggtcatgcatctcctgtgtttcaagatgaaagctctg 332  
Db 374 gcaagctctatttcccaaaaaggctatgatttctctcttcccaagagagagctcag 433  
QY 333 tgcagctctcattgatgcatgcatgattgaagaagatgg-aaacctctacctttgtgtatca 391  
Db 434 ttacagcactcatgatgctgtgtattgaagaagatggaaaaactctacctgtgtgttcc 493  
QY 392 agtccactat----caaggataagccagctccagattcgg--ccttggatctcagtgcac 445  
Db 494 tagccctctactatccaaaggacaaacccagttccaaatcacgtcccttggaaatttaagtgt 553  
QY 446 agtgactttgtggtgtgttcacagccacttgaccacgacgaaaaactatatttgttgt 505  
Db 554 agtgatttgaatggatggtttcagcccttggatcccccgaacaaattttgtgtgga 613  
QY 506 ggtgtctctgacaccattaccagctgtgtgagcttgcggtgtaattggatcggtctatacgg 565  
Db 614 ggtgtctatagccattaaagggtctgtggaactgtctatgatcatgagccggctgttgt 673  
QY 566 ggtgtgtctacgtggattgatcacgacccctgagctaaataaccacaa-aaggagctgg 624  
Db 674 ggaattgtgtatgcaggaattgatcacagatcctgagctaaataaccacaaagctcctgg 733  
QY 625 gagagttgcgtctctcctaacaacagagtttacatagctgctatcagttcccgctttgttca 684  
Db 734 gcgagttgcttctcccaatcagcagagctatattgtgcttattgtcgtgctggtttgttca 793  
QY 685 gctcagcatggagagatagataaacgggttaagccttatactacattttggaataattcta 744  
Db 794 gcttcagcatggtgatattgataaacgtgtggaggttaagccatattgtgctagatgacca 853  
QY 745 gaaatg 750  
Db 854 gatgtg 859

## RESULT 5

ABL09795

ID ABL09795 standard; cDNA; 2748 BP.

AC ABL09795;

XX

XX 26-MAR-2002 (first entry)

DT

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 23867.

DE

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS

XX W0200171042-A2.

PN

XX 27-SEP-2001.

PD

XX 23-MAR-2001; 2001WO-US09231.

PF

XX 23-MAR-2000; 2000US-191637P.

PR

PR 11-JUL-2000; 2000US-0614150.

XX

XX (PEKE ) PE CORP NY.

PA

XX

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB65692.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

PS Claim 1; SEQ ID NO 23867; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 2748 BP; 709 A; 674 C; 694 G; 671 T; 0 other;

## Query Match

Best Local Similarity 7.1%; Score 304.6; DB 23; Length 2748;

Matches 406; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 140 ggggagagagtggaacgatattctcgaaaggtgtttgttagcggtattcctccacagacatt 199

Db 942 ggcgatggcacggctgcgtctctcccggaaggtattcgttggcggtctccacggacatc 1001

QY 200 gatgaagatgagatcacagctagtttgcgtcttggccctctgattgtggtgagctt 259

Db 1002 gatgaggtgagattaccacttcgttctggcggtcggccattggtgctgattggcca 1061

QY 260 cataaagctgagagcaaatcctatttctcctaagactatgcatctcctctgctgtttcaa 319

Db 1062 cacaagggcggaatacgaagtcgtatttccgcgccaaaggatattgcttctctgtctccag 1121

QY 320 gatgaagctctgtgcaggctctcattgatgcattgataagaagatggaacacacttac 379

Db 1122 gacgagagcagtgtagcagcgttaattgactcgtgcatacaggtgaggaagctgtat 1181

QY 380 ctttgtgtatcaagtcaccactatcaagataaagcagctccagattcggccttggaaatc 439

Db 1182 ctatgcgttctctccgacgatcaagataaaggcagtgacagattcctcttggcgcctg 1241

QY 440 agtgacagtgaacttctgtgatggtttcacagccacttgaccacgaaacacatatatt 499

Db 1242 gccgatgcggactatgtcttgatgctaccatgtcactggacccacgcaaacggtgttt 1301

QY 500 gttgtgtgttctctcgaccattacgagctgtgagcttgcgtggttaagtggatggcta 559

Db 1302 gtggcggtgcgtgccacgtctctgaagccttcgaactggcaatgatcatgatgattg 1361

QY 560 tacggaggtgtgctacgctgggattgatccagccctgagctaaataaccacaaagga 619

Db 1362 tacggtggagctgctatgctggaaattgacccctgcggaaattaaagttatcccaaggc 1421

QY 620 gctgggagagttgcgttctcttaataacagagattacatagctgctatcagtgccgcgttt 679

Db 1422 gctggacgttggccttctcgaatcagcagagctacatagcggccatctcagcagattt 1481

QY 680 gttcagctgcagcgtggagagatagataaacgggt 714

Db 1482 gtgcagctgcagcgtggcgtatagacaaagcgggt 1516

## RESULT 6

ABL09879

ABL09879 standard; cDNA; 3523 BP.  
 ABL09879;  
 26-MAR-2002 (first entry)  
 Drosophila melanogaster expressed polynucleotide SEQ ID NO 24119.  
 Drosophila; developmental biology; cell signalling; insecticide;  
 pharmaceutical; gene; ss.  
 Drosophila melanogaster.  
 WO200171042-A2.  
 27-SEP-2001.  
 23-MAR-2001; 2001WO-US09231.  
 23-MAR-2000; 2000US-191637P.  
 11-JUL-2000; 2000US-0614150.  
 (PEKE ) PE CORP NY.  
 Venter JC, Adams M, Li PWD, Myers EW;  
 P-PSDB; ABB65776.  
 WPI; 2001-656860/75.  
 P-PSDB; ABB65776.  
 New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -  
 Claim 1; SEQ ID NO 24119; 21pp + Sequence Listing; English.  
 The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).  
 The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 Sequence 3523 BP; 910 A; 879 C; 911 G; 823 T; 0 other;

Query Match  
 Best Local Similarity 7.1%; Score 304.6; DB 23; Length 3523;  
 Matches 406; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 140 gggagagagtggaacgattctcgaaggtgtttgtagcggtattccctccagacatt 199  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1717 ggcgagtgccaggtcgtctccgcggaaggtattcgttggcgtccaccggacatc 1776  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 200 gatgaagatgagatcacagctagtttcgtcttggccctctgattgtggtggcct 259  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1777 gatgagatgagattaccacttcgttcgcgttcgcgttcggttcgttcgttcgcca 1836  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 260 cataagctgagagcaaacctatttctcctaaaggctatgcattcctctgtttcaa 319  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1837 cacaagcggaatcccaagtcgtatttccgcgaaggtatgcattcctctgttcgag 1896  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 320 gatgaagctctgtcaggtctcattgatgcattgaaggtggaagaaactctac 379  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1897 gacgagcagtgtagcagcagtaattgactcgtgcattcagcaggtgagcaagctgtat 1956  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 380 ctttgtatcaagtccactatcaaggataagccagattccggtcccttggaatc 439  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1957 ctatgcgtttcttcgcgcagcagcaggaaggaagcagtgacagattcgttcgtgcgcgtg 2016  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 440 agtgacagtgaacttgatgagtggttcacagccacttgaccacgaaactatattt 499  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2017 gccgatgcggaactatgtcttgatgctaccatgtcactggaaccacgcaaacggtgttt 2076  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 500 gttgggtgttctcgcaccattacagctgtggagcttgcgatggttaatgatacgacta 559  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2077 gtggggcggtgccacgtctcctgaagccttcgaactcgaatgcatgcatgcatgattg 2136  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 560 tacggaggtgtgtctacgctgggattgatccgaccctgagctaaatacccaaaagga 619  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2137 tacggtggagtatgtctatgtgaattgacaccgatccggaattaaagtatccaaagggc 2196  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 620 gctgggagagttgcgttctcctaacaacagagattacatagctgctatcagtgcccgcttt 679  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2197 gctggagctgtgtggtcttcgaatcagcagagctacatagcggccatctcagccagattt 2256  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 680 gttcagctgcagctgagagatagataaaacgggt 714  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2257 gtgcagctgcagctgagagatagataaaacgggt 2291  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7  
 ABL09813  
 ID ABL09813 standard; cDNA; 3597 BP.  
 XX  
 AC ABL09813;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 23921.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 PI WPI; 2001-656860/75.  
 XX  
 DR P-PSDB; ABB65710.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -  
 XX  
 Claim 1; SEQ ID NO 23921; 21pp + Sequence Listing; English.  
 XX  
 The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).  
 XX  
 The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 Sequence 3523 BP; 910 A; 879 C; 911 G; 823 T; 0 other;





Query Match	7.1%;	Score 304.6;	DB 23;	Length 14372;
Best Local Similarity	70.6%;	Pred. No. 5.4e-47;		
Matches 406;	Conservative 0;	Mismatches 169;	Indels 0;	Gaps 0;
QY 140	ggagagagtgagacgatatctctgaaagggtttgttagcgcggtgctccagacatt	199		
DB 11566	ggcgatggcacggctcgtctcccggaagggtattcgttggtggcgctgccaccgacatc	11625		
QY 200	gatgaagatgagatcacagctagtttttctcgtctttggccctctgattgtgattggcct	259		
DB 11626	gatgagatgagatcacacttctgttccgctcttggccattggcctgctcgatggcca	11685		
QY 260	cataagcctgagagcaaaactatttctctctaaaggctatgcatttctcgtgtttcaa	319		
DB 11686	cacaagcggaatcccaagtcgtatttctccgccaagggtatgccttctcgtgtttccag	11745		
QY 320	gatgaagcctctgacggtctcattgcatgcatgattgaagaagatggaactctac	379		
DB 11746	gacgagacagtgatgcagcagctaatgactcgtgcatcagcagtgagacagctgtat	11805		
QY 380	ctttgtgtatcagctccactatcaaggataagccaagtcagattcggccttggaaatc	439		
DB 11806	ctatgcttcttcgcgcagatcaaggataagcgagtcagattcgtcttggcgccctg	11865		
QY 440	agtgcagctgactttgtggtggtgtctacagccacttgaccacgaaaaactatatt	499		
DB 11866	gcgagtcgcaactatgcttgatgctacatgctcactgacccacgcaaaacgggttt	11925		
QY 500	gttgggtgttctcgcagcattacagactgtgagactgtgagactgtgagatgagcgt	559		
DB 11926	gtggcgcgctgcacgctcctctgaaggccctcgaactgcaatgcatcgtgagattg	11985		
QY 560	tacggaggtgtgtctacgctgggtattgatccgacccctgagctaaaaatacccaagg	619		
DB 11986	tacggtgagatgctatgctggaattgacacccgacatccggaattaaagtatccaaagg	12045		
QY 620	gctggggaggttggttctcttaatacaagatgatacagctgctatcagtgccgcttt	679		
DB 12046	gctggacgtgtggccttctcgaatcagcagagctacatagcggccatctcagccagatt	12105		
QY 680	gttcagctgcagctgagagatagataaaacgggt	714		
DB 12106	gtcgagctgcagctgagctgagatagataaaacgggt	12140		
RESULT 11				
AAV49889/c				
ID	AAV49889 standard; cDNA; 305 BP.			
XX	AAV49889;			
XX	11-NOV-1998 (first entry)			
XX	Human IgA nephropathy-associated gene fragment.			
XX	IgA nephropathy-associated gene; IgA nephropathy diagnosis; therapy; ds.			
XX	Homo sapiens.			
XX	WO9824899-A1.			
XX	11-JUN-1998.			
XX	05-DEC-1997; 97WO-JP04458.			
XX	05-DEC-1996; 96JP-0325763.			
XX	(KYOW ) KYOWA HAKKO KOGYO KK.			
XX	Ishiwata T, Kuga T, Nakagawa S, Nishi T, Nishimura A;			
XX	Sakurada M, Sawada S, Takei M;			
XX	WPI: 1998-333319/29.			



```
DR WPI; 2000-097328/08.
XX DNA sequences preferentially expressed in IgA nephropathy patients,
PT proteins encoded by them, and antibodies to those proteins -
XX
PS Claim 1; Page 140; 180pp; Japanese.
XX
CC This invention describes novel DNA sequences preferentially expressed in
CC IgA nephropathy patients, and DNA sequences stringently hybridizing to
CC them. Independent claims cover diagnostic reagents for IgA nephropathy
CC incorporating the antisense sequences; the treatment of IgA nephropathy
CC using the antisense sequences for mRNA inhibition; proteins associated
CC with IgA nephropathy, containing sequences encoded by the DNA sequences;
CC antibodies recognizing these proteins; the production of the proteins
CC by culture of host cells transformed with DNA encoding them; diagnostic
CC reagents for IgA nephropathy containing the antibodies; and compositions
CC for the treatment of IgA nephropathy which contain the antibodies. The
CC products of the invention can be used for the diagnosis and treatment of
CC IgA nephropathy. This sequence represents a PCR generated human IgA
CC nephropathy-associated protein INP303A cDNA fragment which is described
CC in the method of the invention.
XX
SQ Sequence 305 BP; 95 A; 56 C; 54 G; 94 T; 6 other;

Query Match          7.0%; Score 299; DB 21; Length 305;
Best Local Similarity 98.0%; Pred. No. 3.6e-46;
Matches 299; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2797 acagagttccaggtgttgattataaccacacatgtgtacactataatagaaccacga 2856
Db 305 ACAGAGTTCAGGTTGTGATTATAACCCCAACATGTGTACACTATAAATAGAACACCA 246

QY 2857 gcagggttttccagcagctcagaatcttgtgacgcagtagtcaggcatcttcacaccg 2916
Db 245 GCCAGGCTTTTACGACAGCTCAGAACTCTGTGACGCAGTAGTCAGGCATCTTCACCG 186

QY 2917 acttgaattgaagtgcagttgttggaacttggatcctcttagttgtttttttaa 2976
Db 185 ACTTGAATATTGAAGTGCAGTCTGTGTGGAACTTGGATCATCTAGTTGATTTGTTTAA 126

QY 2977 ttatgattccacatagcaaaaaatccagatccactcaataataatgagggtttatgcta 3036
Db 125 TTATGATTCACATATGACAAAAATCCAGATCCACTCAATTAATAATGAGGTTTATGTCFA 66

QY 3037 tgaataatcctctgtgggtttaatcataaattcttagtctaaacagttggttcactt 3096
Db 65 TGAATAANNNNNGTGGGTTTAAATCTCATAAACATTCTAGTCTAAACAGTTGGCTTCACTT 6

QY 3097 catga 3101
Db 5 CATGA 1

RESULT 13
AAD25036/c
ID RAD25036 standard; DNA; 11003 BP.
AC RAD25036;
XX
XX 12-MAR-2002 (first entry)
XX Human oncostatin M (OSM) gene #1.
XX
XX Human; oncostatin M; OSM gene; haplotyping; genotyping; cancer;
KW lung inflammation; rheumatoid arthritis; chromosome 22q12.2;
XX single nucleotide polymorphism; SNP; ds.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT Variation replace (3536, A)
FT /*tag= a
```

```
FT variation /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (3744, C)
FT /*tag= b
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (4062, A)
FT /*tag= c
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 4087..7003
FT /*tag= d
FT /product= "Human OSM protein"
FT 4087..4120
FT /*tag= e
FT /number= 1
FT 4120..5741
FT /*tag= f
FT replace (5691, G)
FT /*tag= g
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 5742..5884
FT /*tag= h
FT /number= 2
FT 5885..6421
FT /*tag= i
FT replace (5978, T)
FT /*tag= j
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (6354, A)
FT /*tag= k
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (6358, G)
FT /*tag= l
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 6422..7003
FT /*tag= m
FT replace (6458, C)
FT /*tag= n
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (6634, T)
FT /*tag= o
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (6892, A)
FT /*tag= p
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (6995, T)
FT /*tag= q
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (7035, T)
FT /*tag= r
FT /standard_name= "Single nucleotide polymorphism (SNP)"
XX
XX WO200187907-A2.
PN
XX
XX 22-NOV-2001.
XX
XX 17-MAY-2001; 2001WO-US16157.
XX
XX 17-MAY-2000; 2000US-204868P.
XX
XX (GENA-) GENAISSANCE PHARM INC.
XX
XX Duda AE, Kazemi A, Koshy B;
XX
XX WPI; 2002-055680/07.
XX P-PSDB; AAE15318.
XX
XX New isolated human oncostatin M polynucleotide, useful for therapeutic
XX purposes, for studying the expression and function of the
XX polynucleotide and for expressing oncostatin protein -
XX
XX Example 1; Fig 1; 71pp; English.
XX
XX The invention relates to genetic variants of human oncostatin M (OSM)
CC
```



us-09-730-559b-1.rng

```

Qy 2476 ttcattg 2481
    |||||
Db 374 TTCAAG 369

RESULT 15
ABA74962
ID ABA74962 standard; DNA; 225 BP.
XX
XX AC ABA74962;
XX
XX DT 01-FEB-2002 (first entry)
XX
XX DE Human foetal liver single exon nucleic acid probe #23267.
XX
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX OS Homo sapiens.
XX
XX PN W0200157277-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001W0-US00669.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX
XX PR 26-MAY-2000; 2000US-0207456.
XX
XX PR 30-JUN-2000; 2000US-0608408.
XX
XX PR 03-AUG-2000; 2000US-0632366.
XX
XX PR 21-SEP-2000; 2000US-0234687.
XX
XX PR 27-SEP-2000; 2000US-0236359.
XX
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
XX PS Claim 4; SEQ ID NO 23267; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 225 BP; 44 A; 48 C; 66 G; 67 T; 0 other;

Query Match 5.3%; Score 225; DB 22; Length 225;
Best Local Similarity 100.0%; Pred. NO. 1.5e-32;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3353 gtggaagttaagccatagtcttgatgatcagctgtgtgatgaatgtcagggggccgt 3412
Db 1 gtggaagttaagccatagtcttgatgatcagctgtgtgatgaatgtcagggggccgt 60

Qy 3413 tgtggggggaaattgctccattttctgtctaatgttaacctgtgcagttactgt 3472
Db 61 tgtggggggaaattgctccattttctgtctaatgttaacctgtgtgcagttactgt 120

Qy 3473 gaatatgtcggctgctatccattctgtctgtagcagggaattccacagccctggtg 3532
Db 121 gaatatgtcggctgctatccattctgtctgtagcagggaattccacagccctggtg 180

```

Qy 3533 aaggaaggcggtagaccgccctcggcatatttcattccgctggaac 3577  
|||||  
Db 181 aaggaaggcggtagaccgccctcggcatatttcattccgctggaac 225  
|||||

Search completed: September 1, 2002, 09:35:35  
Job time: 9746 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2002, 06:51:40 ; Search time 95.55 seconds  
(without alignments)  
10992.453 Million cell updates/sec

Title: US-09-730-559b-1

Perfect score: 4276

Sequence: 1 ttctaccgttttttcctgc.....atcagaaaaa4276

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2.6/prodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2.6/prodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2.6/prodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2.6/prodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2.6/prodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2.6/prodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	216.8	5.1	17327	1	US-07-906-871-15
2	213.8	5.0	2061	2	US-08-960-022-11
3	212.6	5.0	14747	4	US-09-608-285A-42
4	212.6	5.0	15977	4	US-09-608-285A-59
C 5	210.8	4.9	2713	2	US-08-916-901-6
C 6	210.8	4.9	2713	4	US-09-154-602-6
7	209.8	4.9	2839	4	US-09-061-702-1
8	208.2	4.9	53526	3	US-08-658-136-2
9	208.2	4.9	53577	3	US-08-658-136-1
C 10	207.6	4.9	246240	2	US-08-724-394A-20
C 11	207.6	4.9	246240	2	US-08-724-394A-21
C 12	207.6	4.9	246240	2	US-08-724-394A-22
13	206.8	4.8	1442	2	US-08-454-557C-120
14	206.8	4.8	1442	2	US-08-340-426D-120
15	206.8	4.8	1442	2	US-08-450-673C-120
16	206.6	4.8	246240	2	US-08-724-394A-20
17	206.6	4.8	246240	2	US-08-724-394A-21
18	206.6	4.8	246240	2	US-08-724-394A-22
19	206.4	4.8	1701	4	US-09-078-294-9
20	206.2	4.8	1988	4	US-08-257-963B-11
21	206.2	4.8	1988	4	US-08-367-841A-11
22	206.2	4.8	1988	5	PCT-US95-07201-11
23	206.2	4.8	22481	4	US-08-367-841A-43
24	206.2	4.8	22481	5	PCT-US95-07201-43
25	205.4	4.8	59065	4	US-09-813-817-3
26	205.2	4.8	646	4	US-09-385-982-314
27	203.6	4.8	72928	3	US-09-009-913-1

C 28 202.6 4.7 14636 4 US-09-173-914-6 Sequence 6, Appli  
29 202.2 4.7 31571 1 US-08-323-443B-1 Sequence 1, Appli  
C 30 202 4.7 36741 4 US-09-301-665-3 Sequence 3, Appli  
31 201.8 4.7 619 4 US-09-385-982-358 Sequence 358, App  
C 32 201.6 4.7 2115 1 US-08-395-800A-7 Sequence 7, Appli  
33 200.6 4.7 4823 2 US-08-457-254-5 Sequence 5, Appli  
34 200.6 4.7 4823 2 US-08-484-257-20 Sequence 20, Appli  
35 200.6 4.7 4823 3 US-08-999-927-5 Sequence 5, Appli  
36 200.6 4.7 4823 4 US-08-461-819-5 Sequence 5, Appli  
37 200.6 4.7 4823 5 PCT-US94-08806-28 Sequence 28, Appli  
38 200.6 4.7 4823 5 PCT-US95-01829-5 Sequence 5, Appli  
39 200.6 4.7 4823 5 PCT-US95-16626-5 Sequence 5, Appli  
C 40 200 4.7 2127 1 US-08-832-883-54 Sequence 54, Appli  
C 41 200 4.7 2127 2 US-08-832-877-54 Sequence 54, Appli  
C 42 199.6 4.7 1386 2 US-08-687-080-76 Sequence 76, Appli  
43 199.4 4.7 7676 1 US-08-451-777A-7 Sequence 7, Appli  
44 199.4 4.7 7676 2 US-08-451-778A-7 Sequence 7, Appli  
45 199.4 4.7 7676 2 US-08-998-208-7 Sequence 7, Appli

## ALIGNMENTS

### RESULT 1

US-07-906-871-15/c  
; Sequence 15, Application US/07906871  
; Patent No. 5340739  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; APPLICANT: Avraham, Shalom  
; TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC  
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES  
; TITLE OF INVENTION: THEREOF  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1225 Connecticut Avenue, N.W., Suite 300  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/906,871  
; FILING DATE: 19920103  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/816,289  
; FILING DATE: 03 JAN 1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/635,544  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US89/03051  
; FILING DATE: 13-JUL-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/224,035  
; FILING DATE: 13-JUL-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cimbalis, Michele A.  
; REGISTRATION NUMBER: 33,851  
; REFERENCE/DOCKET NUMBER: 0627.2830004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)833-7533  
; TELEFAX: (202)833-8716  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17327 base pairs  
; TYPE: NUCLEIC ACID

```
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
;   NAME/KEY: exon
;   LOCATION: 621..753
; FEATURE:
;   NAME/KEY: intron
;   LOCATION: 754..9596
; FEATURE:
;   NAME/KEY: exon
;   LOCATION: 9597..9744
; FEATURE:
;   NAME/KEY: intron
;   LOCATION: 9745..16396
; FEATURE:
;   NAME/KEY: exon
;   LOCATION: 16397..17327
; US-07-906-871-15

Query Match          5.1%; Score 216.8; DB 1; Length 17327;
Best Local Similarity 83.9%; Pred. No. 9.3e-36;
Matches 245; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1823 ttgtttgttgagatgagtcctctccgtcacccaggtgagtgagtggttacagtc 1882
Db 5307 TTTTGTGAGACAGAGTCTCTGTCGTCATCCAGGCTGGAGTGCAGTGCATC 5248

QY 1883 tcagctcaactcccaactctgcctcccggttcaagtgtatctcctgccttagccocga 1942
Db 5247 TCAGCTCACTGCAACCTCCACCTCCAGGTTCAAGTGATTCCTCCTGCCCTCAGCCTCCCAA 5188

QY 1943 gtagggtgaattacaggtcatatcacaccatgcccaacaaattgtgtatttttagtgga 2002
Db 5187 GTAGCTGGAATTACAGGTGCACACCACTGCGCCAGTAAATTTTGTATTTTAGTAGAG 5128

QY 2003 acagggtttcacctgtgtgtcaggccagtttcaaaactcctgacctcaagggtaccacct 2062
Db 5127 ACAAGTCTCACCATGTTGCCAGGCTAGTCTCAAACTCTTGACCTCAAGTGATCCACTT 5068

QY 2063 gcctcacccctcaagtgtcgggatacaggcatagcaccaccaccatgcc 2114
Db 5067 GCCTTGCCCTCCCAAAATGCTGGGATTACAGGCATGAGCCACACACCTGGC 5016

RESULT 2
US-08-960-022-11
; Sequence 11, Application US/08960022
; Patent No. 5976837
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,022
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2061 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-960-022-11

Query Match          5.0%; Score 213.8; DB 2; Length 2061;
Best Local Similarity 80.2%; Pred. No. 2.3e-35;
Matches 251; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 1823 ttgtttgttgagatgagtcctctccgtcacccaggtgagtgagtggttacagtc 1882
Db 1230 TTTTGTGAGACAGAGTCTCTGTCACCTCAGGCTGGAGTGCAGTGCACAGTC 1289

QY 1883 tcagctcaactcccaactctgcctcccggttcaagtgtatctccttagccttagcccca 1942
Db 1290 TGGGCTCACTCACTCCTCTGCTACCGGGTTCAAGCAGATTCCTCCTGCTCAGCCTCTCGA 1349

QY 1943 gtagggtgaattacaggtcatatcacaccatgcccaacaaattgtgtatttttagtgga 2002
Db 1350 GTAGCTGGGATTACAGGACCTGTGACCACGCTGGCTAAATTTTGTATTTTAGTTGAG 1409

QY 2003 acagggtttcacctgtgtgtcaggccagtttcaaaactcctgacctcaagggtaccacct 2062
Db 1410 ACGGGGTTTCCACCATGTTGGGCAAGCTAGTCTCGAACTCCTGACCTCAAGTGATCCGCT 1469

QY 2063 gcctcacccctcaagtgtcgggatacaggcatagcaccaccaccatgctggtgacctac 2122
Db 1470 GCCTTGCCCTCCCAAGTGTGGGATTACAGGTGTGAGCCACCGTGCCTGACCTTACCTTT 1529

QY 2123 ttggttttttatg 2135
Db 1530 GAGATTGTGATG 1542
```

```
RESULT 3
US-09-608-285A-42
; Sequence 42, Application US/09608285A
; Patent No. 6335013
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; APPLICANT: Yeung, George
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
; FILE REFERENCE: 28110/36570
; CURRENT APPLICATION NUMBER: US/09/608,285A
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/583,231
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 09/557,800
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481,238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
```

```
; PRIOR APPLICATION NUMBER: 09/350,836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 42
; LENGTH: 14747
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (13641)
; OTHER INFORMATION: n = adenosine or guanine or cytosine or thymidine
US-09-608-285A-42

Query Match          5.0%; Score 212.6; DB 4; Length 14747;
Best Local Similarity 79.7%; Pred. No. 6.5e-35;
Matches 251; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1800 gattagatgttaacatgcctacttctgtttgttgagatggagtctctcgcgtcaacca 1859
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9983 gcttattctctctcccccacccttttttttgagacagatgttgcctgtcaccoca 10042
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1860 ggctggagtgcatgtgtacagctcactcactcaacacctgctcccggttcaagt 1919
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10043 ggctggagtgcatgtgtacagctcactcactcaacacctgctcccggttcaagt 10102
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1920 attctctgccttagcctcccgagtagtggaattacaggatataccaccatgcccaac 1979
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10103 attctctgccttagcctcccgagtagtggaattacaggatataccaccatgcccaac 10162
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1980 aaatgtttgtatttttagtgagacaggggtttaccctgttggtcagccagtttcaaac 2039
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10163 taattttgtatttttagtagagtggtttaccacgttggcaggtgatctcaaac 10222
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 2040 tctgacctcaaggatcacactgcctcaccctcaccctcaagtgctgggattacagccatga 2099
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10223 tctgacctcaaggatcacactgcctcaccctcaccctcaccctcaagtgctgggattacagccatga 10282
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 2100 gccaccaccatgcc 2114
      ||||| | | | |
Db 10283 gccaccgcgcctggc 10297
      ||||| | | | |

RESULT 4
US-09-608-285A-59
; Sequence 59, Application US/09608285A
; Patent No. 6335013
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; APPLICANT: Yeung, George
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: 28110/36570
; CURRENT APPLICATION NUMBER: US/09/608,285A
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/583,231
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 09/557,800
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481,238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 1999-08-09
```

```
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350,836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 59
; LENGTH: 15977
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CD39-L4/L66 Gene Sequence
; NAME/KEY: CDS
; LOCATION: (245)..(461)
; NAME/KEY: CDS
; LOCATION: (1454)..(1533)
; NAME/KEY: CDS
; LOCATION: (2734)..(2877)
; NAME/KEY: CDS
; LOCATION: (4364)..(4439)
; NAME/KEY: CDS
; LOCATION: (4679)..(4714)
; NAME/KEY: CDS
; LOCATION: (5326)..(5414)
; NAME/KEY: CDS
; LOCATION: (5723)..(5802)
; NAME/KEY: CDS
; LOCATION: (6751)..(6812)
; NAME/KEY: CDS
; LOCATION: (7758)..(7859)
; NAME/KEY: CDS
; LOCATION: (8712)..(8852)
; NAME/KEY: CDS
; LOCATION: (9831)..(9887)
; NAME/KEY: CDS
; LOCATION: (11613)..(11728)
; NAME/KEY: CDS
; LOCATION: (13146)..(13691)
; NAME/KEY: CDS
; LOCATION: (15702)..(15839)
; NAME/KEY: misc_feature
; LOCATION: (14871)
; OTHER INFORMATION: n = a or c or g or t
US-09-608-285A-59
```

```
Query Match          5.0%; Score 212.6; DB 4; Length 15977;
Best Local Similarity 79.7%; Pred. No. 6.7e-35;
Matches 251; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1800 gattagatgttaacatgcctacttctgtttgttgagatggagtctctcgcgtcaacca 1859
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 11213 gcttattctctctcccccacccttttttttgagacagatgttgcctgtcaccoca 11272
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1860 ggctggagtgcatgtgtacagctcactcaccctcactcctgcctcccggttcaagt 1919
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 11273 ggctggagtgcatgtgtacagctcactcactcactcactcactcactcactcactcactc 11332
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1920 attctctgccttagcctcccgagtagtggaattacaggatataccaccatgcccaac 1979
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 11333 attctctgccttagcctcccgagtagtggaattacaggatataccaccatgcccaac 11392
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1980 aaatgtttgtatttttagtgagacaggggtttcaccctgttggtcagccagtttcaaac 2039
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 11393 taattttgtatttttagtagagatgggtttcaccacacgttggtccacacgtgatctcaaac 11452
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Qy	2040	tctgacctcaaggga	tccaactgctctcaaccctcaaaagtctgtggattacaggc	atga	2099
Db	11453	tctgacctcaagtgt	tccacccctcagctcccaaaagtctaggattacaggc <td>atga</td> <td>11512</td>	atga	11512
Qy	2100	gccacccaccatgcc	2114		
Db	11513	gccacccgcctggc	11527		

```

RESULT 5
US-08-916-901-6/C
> Sequence 6, Application US/08916901
> Patent No. 5892012
> GENERAL INFORMATION:
> APPLICANT: Hillman, Jennifer L.
> APPLICANT: Lal, Preeti
> APPLICANT: Corley, Neil C.
> APPLICANT: Shah, Purvi
> TITLE OF INVENTION: RAB PROTEINS
> NUMBER OF SEQUENCES: 9
> CORRESPONDENCE ADDRESS:
> ADDRESSSEE: Incyte Pharmaceuticals, Inc.
> STREET: 3174 Porter Dr.
> CITY: Palo Alto
> STATE: CA
> COUNTRY: USA
> ZIP: 94304
> COMPUTER READABLE FORM:
> MEDIUM TYPE: Diskette
> COMPUTER: IBM Compatible
> OPERATING SYSTEM: DOS
> SOFTWARE: FastSeq for Windows Version 2
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/916,901
> FILING DATE: Filed Herewith
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER:
> FILING DATE:
> ATTORNEY/AGENT INFORMATION:
> NAME: Billings, Lucy J.
> REGISTRATION NUMBER: 36,749
> REFERENCE/DOCKET NUMBER: PF-0367 US
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: 415-855-0555
> TELEFAX: 415-845-4166
> INFORMATION FOR SEQ ID NO: 6:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 2713 base pairs
> TYPE: nucleic acid
> STRANDEDNESS: single
> TOPOLOGY: linear
> US-08-916-901-6

```

	Query Match	4.9%	Score 210.8	DB 2	Length 2713
	Best Local Similarity	79.7%	Pred. No. Ie-34		
	Matches 274	Conservative	0	Mismatches 67	Indels 3
				Gaps	
Qy	1811	taacatgctacttgtttgtttgagatggagctctctccgtcacccagcagctggagtgc	1870		
Db	1421	TAAATTTATTTATTTATTTTGTGAGATGGAGTCTCGCTCTGTCAACCCAGGCTGGAGTGC	1362		
Qy	1871	agtgatcacgctcagctcactccaacctctgcctcccggttcaagtgattctcctgcc	1930		
Db	1361	AATGGCATGATCTCGGCTACATGCAACCTCTGCCCTCCGGGTTCAAGTGATTCCTCTGCC	1302		
Qy	1931	ttagcctcccgagtaggtgggaattacaggcatatataccaccatgccccacaaa--tggttg	1988		
Db	1301	TCAGGCTCTGAGTAGCTGGGATTACAGGCATTCGCCACCCGGCCGACGCTAATTTTTTTG	1242		
Qy	1989	LatTTTTagtgagagaca-ggggtttcaccgctgttggtcaggccagttttcaaacctcctgcc	2047		

1241	TATTTTGTAGTAGACAGGGGTTTACCAATGTTGGCCAGGCTGGTCTCGAACTCCTGACC	1181
Db		
2048	tcaaggagatcccaactgcctccatcccccctcaaaagtctcgggattacaaggcatagagccaccca	2107
Qy		
1181	TCAGGTGATCCACACCGCTCGGGCTCCCAAAGTGTGGGATTACAGGCATGAGCCACCGC	1122
Db		
2108	ccatgcctggcctacttgggttttttatgcacactaaaaataacc	2151
Qy		
1121	GCCTGGCCGGACACCACTCTCTTTTCAATGAACAGCTCAATATCC	1078
Db		

```

RESULT 6
US-09-154-602-6/c
: Sequence 6, Application US/09154602
: Patent No. 6300472
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Lal, Preeti
: APPLICANT: Corley, Neil C.
: APPLICANT: Shah, Purvi
: TITLE OF INVENTION: RAB PROTEINS
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESS: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Dr.
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/154,602
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/916,901
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0367 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2713 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-09-154-602-6

```

[illegible]



```
Db 1241 TATTTTATGAGACAGGGGTTTACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACC 1182
QY 2048 tcaaggatccacctccaccctcaaaagtctgggattacaggtacagccaccaca 2107
Db 1181 TAGGTGATCCACCCGCTCGGCTCCCAAAAGTGTGGGATTACAGGATGAGCCACCGC 1122
QY 2108 ceatgctggctacttggttttttatgcacactaaaaataacc 2151
Db 1121 GCCTGCGGACACCACTTCTTTTCAATGAAACAGTCTAATTCC 1078

RESULT 7
US-09-061-702-1
; Sequence 1, Application US/09061702
; Patent No. 6165737
; GENERAL INFORMATION:
; APPLICANT: Wang, Xiaodong
; APPLICANT: Liu, Xuesong
; TITLE OF INVENTION: DNA FRAGMENTATION FACTOR INVOLVED IN
; TITLE OF INVENTION: APOPTOSIS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/061,702
; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McMillian, Nabeela R.
; REGISTRATION NUMBER: P-43,363
; REFERENCE/DOCKET NUMBER: UTSD:546
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512)418-3000
; TELEFAX: (512)474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2839 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-09-061-702-1

Query Match 4.9%; Score 209.8; DB 4; Length 2839;
Best Local Similarity 76.0%; Pred. No. 1.6e-34;
Matches 259; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 1823 ttgtttgttgagatggagttctctccgtccaccagctgagtgagtgacagtc 1882
Db 1984 TTTGTTGTTGATGACAGGGTCTCACTCTGCTCGCTGAGGTGAGTGACATGCAATC 2043
QY 1883 tcagctcaactcccaacctctgcctcccggttcaagtgtattctctctagctccgga 1942
Db 2044 TTGGCTCACTGCAACCTCCGCCCTCTCAGGTTCAGTGGTTCCTCGCTCAGCCTCCCAA 2103
QY 1943 gtagggtgaattacaggtacatataccaccatgcccacaataattgttttagtgagg 2002
Db 2104 GTAGCTGGGATTACAGGCACCTACCACAGGGCAGCTAATTTTGTATGTTAGTAGTA 2163
QY 2003 acaggtttcacggtttgttgagccagtttcaactctcaactcaaggaatccacct 2062
Db 2164 ACGGGGTTTACCATGTTGGCCAGGGCTGTTCTCGAACTCCTGACCTCAGGTGATCGCCC 2223
```

```
QY 2063 gctcaccacctcaaaagtctgggattacaggtacgagccaccaccacatgctgacctac 2122
Db 2224 GCCTCGGTCTCCCAAAGTGTGGGATTACAGGCATGAGCCACTGCACCTGACCTGCTGAA 2283
QY 2123 ttggttttttatgcacactaaaaataaccatcatctcactg 2163
Db 2284 TTGTTTATAATGGCAAGAAATAGGAAACCCCCCAATGCTCTG 2324

RESULT 8
US-08-658-136-2
; Sequence 2, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53526 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-658-136-2

Query Match 4.9%; Score 208.2; DB 3; Length 53526;
Best Local Similarity 75.7%; Pred. No. 7.2e-34;
Matches 258; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 1818 cctactttgtttgttgagatgaggtctctccgtccaccagctgagtgagtgagtgata 1877
Db 989 CTTTCTTTTCTTTTGTGATGGAGTCTCACTCTGTCCACCCAGGCTGGAGTGCAGTGGG 1048
QY 1878 cagtcagtcactccaacctctcccggttcaagtgtattctctcctccttagcct 1937
Db 1049 TAACCTCAGGTACACGCGACCTCCCGGGTTCACCGTTCCTCCTGCCCTCAGCCT 1108
QY 1938 cccgagtgagggaattacaggtacatataccaccatgcccacaataattgttatttttag 1997
Db 1109 CCCGAGTAGTGGGATTACAGGCACCCACCACCCACCGCTGGCTAATTTTGTATTATTAG 1168
```

Qy 1998 tggagacagggtttccaccgtggttgctcaggccaggttccaactcctctgaacctcaaggatc 2057  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1169 TAGAGACGGGGTTTTTCCACAGTTGGCCAGTTGGTCTCGAACTCTTGCCCTCATGTGACC 1228  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 2058 cactgcctcacccccctcaaagtgtcggtattacaggcatgagccacccacacatgcctg 2117  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1229 CGCCTGCCTGGCCTCCCAAGTCTGGGATTACAGGTGTGAGCCACTGTGCCITGGCCTG 1288  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 2118 cctactggtgttttttatgcacacataaaaaaacctacatatc 2158  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1289 GCTTTCCTGTGTTCTTTTCTCTCCTCTTAGTTTCCCCCCTTT 1329  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9  
US-08-658-136-1  
; Sequence 1, Application US/08658136  
; Patent No. 6071717  
; GENERAL INFORMATION:  
; APPLICANT: KLINGER, KATHERINE W  
; APPLICANT: LANDES, GREGORY M  
; APPLICANT: BURN, TIMOTHY C  
; APPLICANT: CONNORS, TIMOTHY D  
; APPLICANT: DACKOWSKI, WILLIAM  
; APPLICANT: GERMINO, GREGORY  
; APPLICANT: QIAN, FENG  
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENZYME CORPORATION  
; STREET: ONE MOUNTAIN ROAD  
; CITY: FRAMINGHAM  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 01701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/658,136  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LASSEN, ELIZABETH  
; REGISTRATION NUMBER: 31,845  
; REFERENCE/DOCKET NUMBER: GEN4-17.8  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 508-872-8400  
; TELEFAX: 508-872-5415  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5357 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-658-136-1







---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2002, 06:17:41 ; Search time 3375.57 Seconds  
(without alignments)  
17097.255 Million cell updates/sec

Title: US-09-730-559B-1  
Perfect score: 4276  
Sequence: 1 ttctaccgtttttccctgc.....atcagaaaaa..... 4276

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	605.4	14.2	1077	11 AK015381	Mus muscu
2	605.4	14.2	1586	11 AK015401	Mus muscu
3	598.2	14.0	860	10 BM456398	AGENCOURT
C 4	507	11.9	529	9 AA772278	AA772278 a142d03.s
C 5	495	11.6	506	9 AI392674	AI392674 tg47c02.x
6	490	11.5	501	10 BE468016	BE468016 hz77c06.x
7	481	11.2	492	10 BE504477	BE504477 hz57a01.x
8	447.6	10.5	1066	11 AK021394	AK021394 Mus muscu
C 9	440.2	10.3	465	10 BE502228	BE502228 hy14e07.x
C 10	423.2	9.9	559	12 AQ883641	HS_5478_B
C 11	420.2	9.8	486	10 BF903366	BF903366 QV1-MT022
C 12	410.6	9.6	425	9 AW237641	xm74h08.x
C 13	408.4	9.6	433	9 AI962924	wt24h06.x
C 14	399.6	9.3	438	10 H71226	ys12e09.s1
15	383.8	9.0	395	9 AI802170	tx25b02.x
16	376	8.8	398	10 H71225	ys12e09.r1
17	363.8	8.5	650	9 BB612073	BB612073 BB612073

18	357.2	8.4	636	10 BE302536	ba67b01.y
19	354.2	8.3	399	10 BM286426	526494 MA
C 20	353.8	8.3	365	9 AI093858	q330405.s
C 21	353	8.3	364	9 AA634469	zu76a06.s
22	345.4	8.1	661	10 BG081288	H3063D07-
C 23	343.2	8.0	465	10 BF092226	PM2-TN009
C 24	337.8	7.9	548	9 AW958536	EST370606
C 25	326.4	7.6	329	10 Z19960	HSAABDCI B
C 26	323.8	7.6	914	10 BI648114	603278431
C 27	320.8	7.5	822	10 BG068252	H3063D07-
C 28	307	7.2	385	10 T93967	YD56d10.s1
29	302	7.1	302	9 AA381126	EST94221
30	299.2	7.0	820	10 BF181733	601805546
31	291.4	6.8	505	10 BE967640	601648932
C 32	290.2	6.8	410	10 H57530	Y705b10.r1
C 33	281.4	6.6	634	10 BE383195	601298746
34	281	6.6	578	10 BI359798	601298746
35	278.8	6.5	523	9 AV282552	384502 MA
C 36	274.2	6.4	521	9 AW653337	101985 MA
37	263.4	6.2	777	12 BHI26248	GG007 Chi
38	258.8	6.1	372	9 AW430013	68737 MAR
C 39	257.2	6.0	438	9 AU022414	AU022414
40	256.8	6.0	661	9 BB405049	BB405049
C 41	252	5.9	321	10 BG956024	PM2-CT080
42	248.6	5.8	593	10 BE977433	bs63h12.y
C 43	245.6	5.7	274	9 AA782211	ai47h12.s
44	239.6	5.6	582	10 BG306825	fm07e09.y
45	236.4	5.5	499	10 BI349383	dac60e03.

## ALIGNMENTS

## RESULT 1

AK015381 1077 bp mRNA linear HTC 19-JAN-2002  
LOCUS Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930444G21:RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain) containing protein, full insert sequence.  
DEFINITION  
ACCSSION AK015381 GI:12853699  
VERSION HTC; CAP trapper.  
KEYWORDS Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library  
SOURCE clone:4930444G21.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (sites)  
AUTHORS Carninci,P. and Hayashizaki,Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2 (sites)  
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
REFERENCE 3 (sites)  
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yanamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)









) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dN) primer  
[5'-TGTACCAATCTGAAGTGGGAGCGCGCACCAATTTTTTTTTTTTTTTTTTTT  
TTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH.

BASE COUNT 196 a 80 c 86 g 167 t  
ORIGIN

Query Match 11.9%; Score 507; DB 9; Length 529;  
Best Local Similarity 98.7%; Pred. No. 8.1e-54;  
Matches 522; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 3740 gtgagattctgtgtaactcaacaatatgaactccttttctgtattgccatcggttgc 3799  
|||||  
Db 529 GTGATCTCTGTGCTACTGCAACAATATGAACCTCTTTTCGTATTGCCATCGGGTTGC 470  
QY 3800 atggaagtattctctctgttttctggaaccagagatcccaacttctctgaacat 3859  
|||||  
Db 469 ATGGAAGTTTATCTCTGTGTTTCTGGAACCAAGAGATCCAAATCTCTGCAACAT 410  
QY 3860 ttcttagaggagagagaaataataaagagaaatgaacaatagagtattttgggtt 3919  
|||||  
Db 409 TTTCTTAGAGGAGAGAGAAATATTAAGAGAAATGAACAATAGAGTATTITGGGTT 350  
QY 3920 tttaataattatgttaataataacataataagaatacttttattataaataaccatg 3979  
|||||  
Db 349 TTTAATAAATTATGTTTAATAATAACATATAAGATCTTTTATTAAATAACCATG 290  
QY 3980 caacataacacatctcgtctctatctgacagtttttcccccgaggagtgctttgccttt 4039  
|||||  
Db 289 CAACAATAACACTATCGGTCTATCTGACAGTTTTCGCCAGGGAAGTGCCTTTGCCCTT 230  
QY 4040 tctcttcttttttttttttttttcac--tttttgtctctctctcttttttccatcccttt 4097  
Db 229 TCTTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 170  
QY 4098 ttaatttttttaacgaatgagagaaatgaacaatttttaataatgaagacatgttaga 4157  
Db 169 TTAATTTTTTAAACGAATGGGAAGTTAACAAATTTTAAATGGAAGACATGTTAGA 110  
QY 4158 gcaacaaatgcataagcaagactgagcagcattataataattttcagggttttgaggc 4217  
Db 109 GCAACAATAATGCATAAGCAAGACTGAGCAGCATTTATAATTTTTCAGGCTTTTGAGGC 50  
QY 4218 tgaacataatttcattatccctcacaaggtttaccaccacacagaaaaa 4266  
Db 49 TGAACATAATTTTCATTATCCCTCAAAAAAGTTACCACCACATCAGAAAAA 1

RESULT 5  
LOCUS AI392674 506 bp mRNA linear EST 04-FEB-1999  
DEFINITION t947c02.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:2111906 3', mRNA sequence.  
ACCESSION AI392674  
VERSION AI392674.1 GI:4222221  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 506)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL  
COMMENT

Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from Gibco  
High quality sequence stop: 444.  
Location/Qualifiers

FEATURES  
source

1. 506  
/organism="Homo sapiens"  
/db\_xref="taxon.9606"  
/clone="IMAGE:2111906"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NbH19W, testis NHT, and B-cell  
NCI-CGAP-GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
728408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 185 a 75 c 82 g 164 t  
ORIGIN

Query Match 11.6%; Score 495; DB 9; Length 506;  
Best Local Similarity 99.8%; Pred. No. 2.5e-52;  
Matches 506; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 3762 acaatatgaactccttttctgtattgccatcggttgcagtggaagttttattctctgtt 3821  
Db 506 ACAATATGAACCTCTTTTCGTATTGCCATCGGGTTCATGGAGATTTTATTCCTGTGTT 447  
QY 3822 ttgctgaaaccaagagatccaaactctctgcaacatttcttagaggagagagaaa 3881  
Db 446 TTGCTGGAACCAAGAGAGATCCAAACTTCTGCAACATTTCTTAGAGGAGAGAGAGAA 387  
QY 3882 tattaaagagaatgaacaatagagtatttgggtttttaattaaattattgttaata 3941  
Db 386 TATTAAAGAGAAATGAACAATAGAGTATTTTGGGTTTTTAATTAATTAATTTGTTAATA 327  
QY 3942 atataacataaagaaacttttattataaataaccatgcaacaataacacatcgtctta 4001  
Db 326 ATATAACATATAAGATACCTTTTATTAAATACCATGCAACAATAACACTATCGGTCTA 267  
QY 4002 tctgacagtttttccccagggagtgcttttgcctttctcttcttctttttttttttttt 4061  
Db 266 TCTGACAGTTTTTCCCCCAGGGAAGTCTTTTGCCTTTTCTTTCTTTTCTTTTCTTTT 207  
QY 4062 catctttttgtctctctcttttttccatcccttttttaatttttttaacagcaatggag 4121  
Db 206 CATCTTTTGTG-TCTCTCTCTTTTTCATCCCTTTTAAATTTTAAACAGCAATGGAG 148  
QY 4122 gaagtttaacaatttttaaggaagagcatgttagagcaacaataatgcataaagcaagact 4181  
Db 147 GAAGTTAAACAATTTTAAATGGAAGAGCATGTTAGACAAACAATAATGACATAGCAAGACT 88  
QY 4182 gacgacattataataattttcagggttttgaggctgacaataatttcattatccctca 4241  
Db 87 GAGCAGCATTTATAATTAATTTTCAGGGTTTTTGAGGCTGAACATAAATTCATTATTCCTCA 28

QY 4242 aaagttaccacacacacagaaaaaa 4268  
Db 27 AAAAGTTACCACCACATCAGAAAAAA 1

RESULT 6



sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 11076861  
 4 (sites)  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 585-590 (2001)  
 5 (bases 1 to 1066)  
 Adachi.J., Aizawa.K., Akahira.S., Akimura.T., Aono.H., Arai,A.,  
 Arakawa,T., Baidarelli.R., Bono.H., Brownstein,M., Bult,C.,  
 Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,  
 Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,  
 Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa.M., Kasukawa,T.,  
 Kato,H., Kawai,J., Kojima.Y., Konno.H., Kouda.M., Koya.S.,  
 Kurthara.C., Matsuyama.T., Miyazaki.A., Nishi.K., Nomura,K.,  
 Numazaki.R., Ohno.M., Okazaki.Y., Okido,T., Owa.C., Quackenbush,J.,  
 Saito,H., Saito.R., Sakai.C., Sakai.K., Sano,H., Sasaki.D.,  
 Schriml,L., Shibata,K., Shibata.Y., Shinagawa.A., Shiraki,T.,  
 Sogabe.Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,  
 Tanaka.T., Tejima,Y., Toya.T., Yamamura.T., Yamanaka.I.,  
 Yasunishi.A., Yoshida.K., Yoshino.M., Muramatsu,M. and  
 Hayashizaki,Y.  
 Direct Submission  
 Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
 URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
 Fax:81-45-503-9216)  
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
 further details.  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues. First strand cDNA was primed with a primer  
 [5'-GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTTN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse transcriptase  
 and subsequently enriched for full-length by cap-trapper. cDNA went  
 through one round of normalization to Rot = 10.0 and subtraction to  
 Rot = 239.4. Second strand cDNA was prepared with the primer  
 adapter of sequence [5'  
 GAGAGAGAGATCTCGATTGAATTAATTCCTCCCTCCCTCC 3']. cDNA was cleaved  
 with BamHI and XhoI. Vector: a modified pBluescript KS(+) after  
 bulk excision from Lambda FLC I. Cloning sites, 5' end: Sall; 3'  
 end: BamHI. Host: DH10B.  
 Location/Qualifiers  
 1..1066  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="MGD:MG1:1899445"  
 /db\_xref="taxon:10090"  
 /clone="E130118616"  
 /tissue.type="eyeball"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="0 day neonate"  
 <1..427  
 /note="RNA recognition motif. (a.k.a. RRM, RBD, or RNP  
 domain) containing protein  
 data source:Pfam, source key:PF00076, evidence:ISS  
 putative"  
 /codon\_start=2  
 /protein\_id="BAB322395.1"  
 /db\_xref="GI:12862272"  
 /translation="RPLRAVELAMIMDRLYGGVYAGIDTDPKLYPKGAGVAFPSNQ  
 QSVIAISARFVQLQGEIDRKVEKPYVLDLQDCQGACGCKGKFAFFCANNVCL  
 QYTCYCAAIHSRAGREHKPLVDKGGDRPHISFRWN"  
 291 a 221 c 238 g 316 t

Query Match		10.5%;	Score 447.6;	DB 11;	Length 1066;
Best Local Similarity		79.7%;	Pred. No. 1.2e-46;		
Matches 719;		Conservative 0;	Mismatches 129;	Indels 54;	Gaps 14;
QY	3350	cagggtgaagttaagccaatgtcttgatgcacgtgtgtgatgaatgcaggggcc	3409		
Db	197	CGGTGGAGGTTAAGCCATATGCTTGGATGACACAGCTGTGTGATGAATGTCAGGGGCC	256		
QY	3410	cgttggtggggaaattgtccattttctgtgctaaatgtttacctgtctgcagtattac	3469		
Db	257	CGTTGTGGGGGAAATTTGCTCCATTTTCTGTGCTAAATGTTACCTGTCTGCAGTATTAC	316		
QY	3470	tgtgaattctgctgggctgtcattctctgtgctggcagggaattccacaagccctg	3529		
Db	317	TGTGANATATTGCTGGGCTGCTATTACTCTCGTGTGCAGAGAAATCCACAGCCCTTG	376		
QY	3530	gtgaaggaaggcgtgacccctcgccatatttcattccctgcgtgaaact----	3585		
Db	377	GTGAAGGAAGGTGGTACCCTCGGCATATTTTCATTCGCTGGAACCTGAAGAATGAAT	436		
QY	3586	aactgcagtgtcattttcaggcctcagaataaagtcacactcttctgttcattctgcaccc	3645		
Db	437	GACTCCAGAGCTCATTTGACGGCTCAGAGAAGTGCCTCTTTTGTAAATTC-CAACTC	495		
QY	3646	ttctcaacctcttcacgtggtgccttctttttagcagctgtgaacttaactatagta	3705		
Db	496	TTCTCTAGCCTCTTCACACTGGCATATTTTCATTCGCTGGAAC-TATCAGTGGTA	553		
QY	3706	taatgaagaagaatgacactataataggtgtttgttagattctgtgacgtgcaacaa	3765		
Db	554	GAATGAAGAAGAACCACTTA-CACATAGGTATTTTGTAGACTTCGTGTCACTGCAACAG	612		
QY	3766	tatgaactcct--ttttcgtattgccatcgtgtgcatgggaagtttt-attctctgttt	3822		
Db	613	TATGTTCAACTCCATTTCACATTTGCCATTATGTTGCATGCAGGTTTAAATTCCTCTT	672		
QY	3823	tgtcggaaaccaagagatccaaactcctgcgaacatttctcttagaggagagagaaat	3882		
Db	673	TGCTGAGGCGCAAGAGGATCTAACTTGCTGTAGCATTTTCTTAGAG-----GAGAGAAAT	728		
QY	3883	attaaagagaaatgaacaaatagagtattttgggttttaataattattgtttaataa	3942		
Db	729	ATTAACAGAGAAATGAGACAATA-----TTTGTAGGTTTAAATGCTGGTGAATGA	778		
QY	3943	tataacataagaatacatttttataaataaaccatgaacacataaacactatcggtctat	4002		
Db	779	TATAACATATAAGAAACTTTATTGAGATAATC--ACCAACAGTAACACTATCATCTGTT	836		
QY	4003	ctgacagttttccccccagggaagtcgttttgccttttctcttcttcttttttttttc	4062		
Db	837	CTGACACAGTCTCCCCAGGAAGTGTCTTGCAGTCTTTTCTT-----882			
QY	4063	atctttttgtctctcttcttttccatcccttttttaatttttttaacagcaatggagg	4122		
Db	883	---TCATTCTCTCTCTCTCTTTTCTATCCCTTTTA--TTCCTTAACAGCAATGGAGG	937		
QY	4123	aagtttaacaaatttttaa-tggaaagagcatgttagagcaacaaatgcataaagcaagact	4181		
Db	938	AAGTTAACAGTATTATAACAGGAAGAGTATGTCAGAGCAAGCAAAATGCATGACCAAGATT	997		
QY	4182	gagcagcattataataattttcaggggttttgagcgtgaacataaatttcattatccctca	4241		
Db	998	GAGCAG-----TGTAACCATTAAGAGCTTTTGAGGCTGATGATAGCTTGTATGCCCTCA	1050		
QY	4242	aa 4243			
Db	1051	AA 1052			

RESULT 9  
BE502228/c  
LOCUS BE502228 465 bp mRNA linear EST 04-AUG-2000

DEFINITION	hvl14e07.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3197316 3', mRNA sequence.
ACCESSION	BE502228
VERSION	BE502228.1 GI:9704636
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1. (Bases 1 to 465)
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaaps-r@mail.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LENL, send email to: info@image.llnl.gov Seq primer: -40UP from Gibco High quality sequence stop: 455. Location/Qualifiers 1. 465 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3197316" /clone_lib="NCI_CGAP_GC6" /tissue_type="pooled germ cell tumors" /lab_host="DH10B" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1469084-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."
FEATURES	174 a 69 c 74 g 148 t
BASE COUNT	
ORIGIN	
Query Match	10.3%; Score 440.2; DB 10; Length 465;
Best Local Similarity	98.1%; Pred. No. 1.6e-45;
Matches 456;	Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY	3798	gcattggaagtttattctctgttttgcggaacagaggatccaaacttcttccaac	3857
Db	465	GCATGGAAGTTTATTCTTGTTTCTGTAACCAAGAGGATCCAAACTTCTTGCAC	406
QY	3858	atttcttagagagagagaaattataaagagaaatgaacaatagagtagattttggg	3917
Db	405	ATTTTCTTAGAGGAGAGAGAAATATTAAGAGAAATGAACAATAGAGTATTTTGGG	346
QY	3918	tttttaataaattattgttaataataacataataaacttttttataaataacca	3977
Db	345	TTTTTAATTAATTTGTTAATAATAACATATAAGAAATAGTTTTTATAAATAACCA	286
QY	3978	tgcaacaataacactatcggtctatctgcaggttttccccccaggaagcttttgccct	4037
Db	285	TGCAACAATAACACTATCGCTCTATCTGACAGTTTTTCCCCCAGGGAAGTCTTTGCCCT	226
QY	4038	tttcttcttttttttttttttttttctct- ttttgttctctcttttttttccctct	4096
Db	225	TTTCCTTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCCATCCCTT	166
QY	4097	tttaatttttttaacagcaatggaggaagttaacaatttttaatggaaagagcatgttag	4156

Db	165	TTTTAATTTTTTTTAAACAGCAATGAGGAAGTTACAAATTTTATGTAAGACGATCTTAG	106
QY	4157	agcaaacaaatgcataagcaagactgagcagcattataataatttcagggttttgagg	4216
Db	105	AGCAAAACAATGCATAAGCAAGACTGAGCAGCATTTATAATTAATTTTCAGGGTTTGAGG	46
QY	4217	ctgaacataatttcattatccctcctcaaaaagtaccaccacatcag	4261
Db	45	CTGAACATAATTTTCATTATCCCTCAAAAAGTTACCACCACATCAG	1
RESULT	10		
AQ883641/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	2997	aaaatccagatccactaatataaatgaggtttatgtctatgaataatctctgggtt	3056
Db	551	AAAACCTCAGACTCACTAAT-AAAATGAGTTTATGTCATATG-ATATTCTCGTGGGTT	495

```

/clone_lib="WT0228"
/dev_stage="Adult"
/notes="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORREST PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
131 a 129 c 97 g 129 t
BASE COUNT
ORIGIN

```

Query Match	9.83;	Score 420.2;	DB 10;	Length 486;
Best Local Similarity	97.1%;	Pred. No. 4.6e+43;		
Matches 428;	Conservative 0;	Mismatches 13;	Indels 0;	Gaps 0;

  

QY	292	taaaggctatgcattccctgcgtgtttc	caagatgaaagctctgtgcagcgtctcatigatgc	351
DB	486	TAAAGGCTATGATTCCTGCTGTTT	CAAGATGAAGCTCTGTGCAGGCTCTCATTTGATGC	427
QY	352	atgcattgaagaagatgaaacctacc	cttctgtgtatcaagtcaccactatcaagataa	411
DB	426	ATGCATTTGAAGAAGATGAAAAC	TCTACCTTTGTGTATCAAGTCCCACTATCAAGGATAA	367
QY	412	gccagtcacagattcggccttggaa	cttcagtcacagtgactttgtgatggatgggttcaca	471
DB	366	GCCAGTCCAGATTCCGCCCTTGG	AATCTCAGTGACAGTGACTTTGTGATGGATGGTTTCACA	307
QY	472	gccacttgaccacgaaatactata	tatttggtggttcctcgaccattacgagctgt	531
DB	306	GCACCTTGACCCACGAAAACAT	ATATTTGTGGTGTGTTCCTCGACCACTATCAGAGCTGT	247
QY	532	ggagcttgcgattggaattgattgc	gctatcacgaggtgtgtcagctgggattgatac	591
DB	246	GGATCTTGGATGATAATGGATCG	GCTATATGGAGGTGTGCTACGCTGGGATTTGATAC	187
QY	592	cgaacctgagctaaatacccaaa	gagctgggaggttgcttctctaatcaacagag	651
DB	186	CGACCCCTGAGCTTAAATAATAC	CCAAAGAGCTGGGAGAGTTGCGTTCTCTTAATCAACAGAG	127
QY	652	ttacatagctgctatcagtcgcc	ctttgttcagctgcagcatggagagatagataaacg	711
DB	126	TTACATAGCTGCTATCAGTGC	CCGCCCTTTGTTACGTGTCAGCTGTCAGATGGAGATAGATAAACG	67
QY	712	ggtaagccttatactacattt	732	
DB	66	GGTGAAGTTAAGCCATATGT	46	

RESULT	12
AW237641/c	
LOCUS	
DEFINITION	linear EST 13-DEC-1999 xm74h08.xl NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2689983 3', mRNA sequence.
ACCESSION	AW237641
VERSION	AW237641.1
KEYWORDS	GI:6570030
SOURCE	EST.
ORGANISM	human. Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 425)
TITLE	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapos-femail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmett-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 352.

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

```

1. 425
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:2689983"
    /clone_lib="NCI_CGAP_Kid11"
    /lab_host="DH10B"

```

/a/mot= DR105  
/note=Organ: Kidney; Vector: pT73D-Pac. (Pharmacia) with a modified polylinker; Site1: Not I; Site2: Eco RI; Plasmid DNA from the normalized library NCI\_CGAP\_K13 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502853). Subtraction by Bento Soares and M.

BASE COUNT	156 a	58 c	66 g	145 t
ORIGIN	FACILMA BONDADO.			

Query Match	9.6%;	Score 410.6;	DB 9;	Length 425;
Best Local Similarity	97.9%;	Pred. No. 7.6e-42;		

QY	3845	aactctcgaacattttcttagagagagagaaataattaaaagagaaatgaacaat	3904
Db	425	AACTTCCTGCAACATTTTCTTAGAGGAGAGAGAAATATTTAAAGAGAAATGAACAAT	366
QY	3905	agagtatattgggtttttaattaaattatgttaataataataacataaagaataactttt	3964
Db	365	AGAGTATTTTGGGTTTTTAATAATTATTGTTAATAATATACAATAAGAATACATTTT	306
QY	3965	atlaaaataaccatgcacaataaacacatacgggtctctatctgcagtttttccccccaggga	4024
Db	305	ATTAAATAAACCATGCAACAATAAACACTATCGGTCTATCTGACAGTTTTTTCCCCCAGGA	246
QY	4025	agtgcattttgccttttcccttttcttttttttttttttttttttttttttttctctctctctt	4084
Db	245	AGTGCATTTGCCCTTTTCCTTTTCCTTTTTTTTTTTTTTTTTTTTTTTTTCATCTTTTGTCTCTCTTT	186

[illegible]

RESULT 13  
AI962924/C

LOCUS AI962924 433 bp mRNA linear EST 09-MAR-2000  
DEFINITION wc24h06.x1 NCI\_CGAP\_Utl1 Homo sapiens cDNA clone IMAGE:2508443 3',  
mRNA sequence.  
ACCESSION AI962924  
VERSION AI962924.1 GI:5755637  
KEYWORDS EST,  
SOURCE human.



ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 433)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: crapsb-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 752 Std Error: 0.00  
Seq primer: -40up from Gibco  
High quality sequence stop: 362.  
FEATURES  
source  
1. .433  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2508443"  
/clone\_lib="NCI CGAP Ut1"  
/tissue\_type="well-differentiated endometrial  
adenocarcinoma, 7 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 1.75 kb. Life Technologies catalog #:  
11538-014"  
BASE COUNT 153 a 59 c 68 g 151 t 2 others  
ORIGIN  
Query Match 9.6%; Score 408.4; DB 9; Length 433;  
Best Local Similarity 97.9%; Pred. No. 1.4e-41;  
Matches 423; Conservative 0; Mismatches 8; Indels 1; Gaps 1;  
QY 3845 aacttcctgcaacattttcttagagagagagagaataattataaagaagaatgaacaat 3904  
Db 433 AACTTCTCTCAACATTTCTTAGAGGAGAGAGAGAAATATTAAGAGAGAAATGAACAAT 374  
QY 3905 agagatttttggttttaattaaattattgttaataataataaataaataactttt 3964  
Db 373 AGAGTATTTTGGGTTTTAAATTAATTTATTTGTAATAATAAATGAAGTACTTTT 314  
QY 3965 attaaataaccatgaacaataacactatcggtctatctgacagttttccccaggga 4024  
Db 313 ATTAAATTAACCATGCAACATTAACATATCGGTCTATTGACAGTTTTCCTCCAGGGA 254  
QY 4025 agtgccttttgcttttcctctctttttttttttttttttttttttttttttttttt 4084  
Db 253 AGTGCCTTTTGCTTTTCCCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 195  
QY 4085 ttctcatcccttttaattt 4144  
Db 194 TTTCCATCCCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 135  
QY 4145 agagcatttagagcaacaataatgataagcaagctgagcagcattataatttttc 4204  
Db 134 AGAGCATGTTAGAGCAACAAATGCAATAGCAAGACTGAGCAGCATTAATTAATTTTC 75  
QY 4205 agggttttgaggtgaacataattttcattatccctcaaaaaagttaccaccacagaaa 4264  
Db 74 AGGGTTTGGGCTGAACATAATTTTATTATCCCTCAAAAAAGTTTCCCCCCCATCAGAAA 15  
QY 4265 aaaaaaataaaa 4276  
Db 14 AAAAAAAAAA 3

RESULT 14  
H71226/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

H71226 438 bp mRNA linear EST 26-OCT-1995  
Ys12e09.s1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone  
IMAGE:214600 3', mRNA sequence.  
H71226  
H71226.1 GI:1043042  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 438)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissio, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins  
M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore  
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,  
Schellenberg, K., Soares, M.B., Tan, F., Thierri-Meg, J., Trevisan, E.,  
Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1184  
High quality sequence stops: 372  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 1184 Std Error: 0.00  
Seq primer: Promega -2lml3  
High quality sequence stop: 372.  
Location/Qualifiers  
1. .438  
/organism="Homo sapiens"  
/db\_xref="GDB:3779504"  
/db\_xref="taxon:9606"  
/clone="IMAGE:214600"  
/clone\_lib="Soares fetal liver spleen lNFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' AACGTGAGAGATTAATAAGATCTTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

FEATURES  
source

BASE COUNT 135 a 90 c 98 g 112 t 3 others  
ORIGIN  
Query Match 9.3%; Score 399.6; DB 10; Length 438;  
Best Local Similarity 96.1%; Pred. No. 1.7e-40;  
Matches 419; Conservative 0; Mismatches 16; Indels 1; Gaps 1;  
QY 3444 ctaatgttaactgtctgacttactgtg-aatatgtgggctgtatccattctcgt 3502  
Db 436 CTAGGTTTACCTGCTGCTGCAATATTACTGTGAATATTCTGGGCTGTACCAATCTCGT 377  
QY 3503 gctggcagggaatcccaagccccctggtgaaggaagcggtgaccgcctcgcatatt 3562  
Db 376 GCTGGCAGGGAATTCCCAACAGCCCGGTGAAGGNANAGGTGAACCCCTCGGCATATT 317

```
Qy 3563 tcattccgtggaactaaagataaactgcagtgctcctatttccaggccctcagaataagtc 3622
Db 316 TCATTCCGCTGGAACCTAAAGGATAAATGAGTGCAGTCTCATTTTCAGGCCTCAGATAAGTGC 257
Qy 3623 actctctgttcattctgaccccttcctcaacctctctcaagctggtgcatgctcctttttag 3682
Db 256 ACTCTCTGTGTCATTCCTGACCCCTTCCTCAACCTCTTTCACGCTGGCATGTCCTTTTGTAG 197
Qy 3683 cagctgttaacttaactatagataaagaagaatgacccataataatagtggttttttga 3742
Db 196 CAGTCTGTAACCTATAGTATATAGTAAGAAAGAAAGAACCTATATATAGTGTTTTGTGA 137
Qy 3743 gattctgttgcactgcaacaataatgaactccttttctgattgcatcggtgttgcattg 3802
Db 136 GATTCTTGTGTCTACGCTGCAACCAATATGAACCTCTTTTTCGATTGCAATCGGGTTGCAATG 77
Qy 3803 gaagcttttattctctgttctgctggaacccaagagagatcccaacttcctgcaacatttt 3862
Db 76 GAAGTTTATTCCTCTGTTTGTGGAACCAAGAGGATCCAAACTTCCTGCAACATTTT 17
Qy 3863 cttagaggagagagag 3878
Db 16 CTTAGAGGAGAGAG 1

RESULT 15
AI802170
LOCUS
DEFINITION
tx25b02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2270571 3',
mRNA sequence.
ACCESSION
AI802170
VERSION
AI802170.1 GI:5367642
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 395)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Cloned by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 462 Std Error: 0.00
Seq primer: -400P from Gibco.
Location/Qualifiers
1..395
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2270571"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI_CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (ClonesIDs
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."
```

BASE COUNT  
ORIGIN

Query Match 9.0%; Score 383.8; DB 9; Length 395;  
Best Local Similarity 98.2%; Pred. NO. 1.6e-38;  
Matches 388; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy 3827 ggaacccaagagatcccaaaccttcctgcaacatttcttagaggagagagaaatatta 3886
Db 1 GGAACCAAGAGATCCCAACCTTCCTGCAACATTTCTTAGAGGAGAGAGAAATATTA 60
Qy 3887 aaagagaatgaacaatagagattttgggtttttaataattattgttaataatata 3946
Db 61 AAAGAGAAATGAACAATAGAGTATTTTGGGTTTAAATTAATTTAAATTAATATA 120
Qy 3947 acatataagaataactttttattataaataaccatgaacaataacactatcgtctatctga 4006
Db 121 ACATATAAGAATACATTTTATTAAAAATAACCATGCAACAATAACACTATCGTCTATCTGA 180
Qy 4007 cagtttttccccccaggggaagtgccttttgccttttcttcttcttttttttttttctatct 4066
Db 181 CAGTTTTTCCCCCAGGGAAGTGCCTTTTCCCATCCCTTTTAAATTTTAAACAGCAATGGAGGAAGT 300
Qy 4067 ttttggctctctctcttttttccatccctttttaaatttttaacagcaatggagggaagt 4126
Db 241 TTTTGTGCTCTCTCTCTTTTCCCATCCCTTTTAAATTTTAAACAGCAATGGAGGAAGT 300
Qy 4127 taacaatttttaaggaagagcatgttagagcaacaacaatgcataagcaagactgagca 4186
Db 301 TAACAATTTTAAATGGAAGAGCATGTTAGAGCAACAATAATGCATAAGCAAGACTGAGCA 360
Qy 4187 gcattataattaattttcagggttttgaggctgaa 4221
Db 361 GCATTATAATTAATTTTCAGGGTTTTCAGGGCTGAA 395
```

Search completed: September 1, 2002, 09:22:57  
Job time: 11116 sec

---





Percent Similarity: 100.000 Percent Identity: 85.714

## alignment\_block:

US-09-730-559B-45 x AA12174 ..

Align seg 1/1 to: AA12174 from: 1 to: 95

1 GGGCTTAATATTATTCATAGA 21  
|||||:|||||:|||||  
68 GlyLeuAsnIleValHisArg 74

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1999.DAT:AA1999

## seq\_documentation\_block:

ID AA1999 standard; Protein: 227 AA.

XX

AC AA1999;

DT 06-DEC-1999 (first entry)

XX Human foetal kidney secreted protein pm633\_1.  
DE Secreted protein; pm633\_1; human; therapy; diagnosis; vaccine;  
KW kidney.  
XX

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 28..40

FT Protein /note= "signal peptide"

FT Protein 41..227

FT Protein /note= "mature protein"

XX WO9947555-A1.

PN

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Percent Similarity: 100.000 Percent Identity: 85.714

## alignment\_block:

US-09-730-559B-45 x AA12174 ..

Align seg 1/1 to: AA12174 from: 1 to: 95

1 GGGCTTAATATTATTCATAGA 21  
|||||:|||||:|||||  
68 GlyLeuAsnIleValHisArg 74

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1999.DAT:AA1999

## seq\_documentation\_block:

ID AA1999 standard; Protein: 227 AA.

XX

AC AA1999;

DT 06-DEC-1999 (first entry)

XX Human foetal kidney secreted protein pm633\_1.  
DE Secreted protein; pm633\_1; human; therapy; diagnosis; vaccine;  
KW kidney.  
XX

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 28..40

FT Protein /note= "signal peptide"

FT Protein 41..227

FT Protein /note= "mature protein"

XX WO9947555-A1.

PN

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

## alignment\_scores:

Quality: 36.00 Length: 7

Ratio: 5.143 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 85.714

## alignment\_block:

US-09-730-559B-45 x AA1999 ..

Align seg 1/1 to: AA1999 from: 1 to: 227

1 GGGCTTAATATTATTCATAGA 21  
|||||:|||||:|||||  
68 GlyLeuAsnIleValHisArg 74

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT:AA2001

## seq\_documentation\_block:

ID AA2001 standard; Protein: 360 AA.

XX

AC AA2001;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

## alignment\_scores:

Quality: 36.00 Length: 7

Ratio: 5.143 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 85.714

## alignment\_block:

US-09-730-559B-45 x AA1999 ..

Align seg 1/1 to: AA1999 from: 1 to: 227

1 GGGCTTAATATTATTCATAGA 21  
|||||:|||||:|||||  
68 GlyLeuAsnIleValHisArg 74

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT:AA2001

## seq\_documentation\_block:

ID AA2001 standard; Protein: 360 AA.

XX

AC AA2001;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

## alignment\_scores:

Quality: 36.00 Length: 7

Ratio: 5.143 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 85.714

## alignment\_block:

US-09-730-559B-45 x AA1999 ..

Align seg 1/1 to: AA1999 from: 1 to: 227

1 GGGCTTAATATTATTCATAGA 21  
|||||:|||||:|||||  
68 GlyLeuAsnIleValHisArg 74

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT:AA2001

## seq\_documentation\_block:

ID AA2001 standard; Protein: 360 AA.

XX

AC AA2001;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

## alignment\_scores:

Quality: 36.00 Length: 7

Ratio: 5.143 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 85.714

## alignment\_block:

US-09-730-559B-45 x AA1999 ..

Align seg 1/1 to: AA1999 from: 1 to: 227

1 GGGCTTAATATTATTCATAGA 21  
|||||:|||||:|||||  
68 GlyLeuAsnIleValHisArg 74

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT:AA2001

## seq\_documentation\_block:

ID AA2001 standard; Protein: 360 AA.

XX

AC AA2001;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 360 AA;

alignment\_scores:  
Quality: 35.00 Length: 7  
Ratio: 5.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 71.429

alignment\_block:  
US-09-730-559B-45 x AAU61249 ..

Align seg 1/1 to: AAU61249 from: 1 to: 360

1 GGGCTTAATATTATTCATAGA 21  
|||||:|||||  
30 GlyLeuAsnValValHisArg 36

seq\_name: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2001.DAT:ABB68264

seq\_documentation\_block:  
ID ABB68264 standard; Protein; 788 AA.

XX  
AC ABB68264;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 31584.

XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
KW  
XX Drosophila melanogaster.

OS  
XX WO200171042-A2.  
PN  
XX 27-SEP-2001.

XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

PI  
XX WPI: 2001-656860/75.  
DR  
XX N-PSDB; ABL12367.

XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

XX Disclosure; SEQ ID NO 31584; 21pp + Sequence Listing; English.

XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABBS7737-ABBS7207).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 788 AA;  
alignment\_scores:  
Quality: 35.00 Length: 8  
Ratio: 4.375 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:  
US-09-730-559B-45 x ABB68264 ..

Align seg 1/1 to: ABB68264 from: 1 to: 788

1 GGGCTTAATATTATTCATAGATCG 24  
|||||:|||||:|||||  
118 GlyLeuAsnIleIleHisLysAla 125

seq\_name: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1997.DAT:AAW27851

seq\_documentation\_block:  
ID AAW27851 standard; Protein; 115 AA.

XX  
AC AAW27851;

XX 21-JUL-1998 (first entry)

XX Staphylococcus aureus protein of unknown function.

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;  
KW Staphylococcal gene; regulatory element; bacterial gene expression;  
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;  
KW toxic shock syndrome.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

FT Misc-difference 1..115 /note= "residues designated X are not defined in  
FT the specification"

FT Misc-difference 1 /note= "Met encoded by CTG"

XX WO9730070-A1.

XX 21-AUG-1997.

XX 19-FEB-1997; 97WO-US02318.

XX 20-FEB-1996; 96US-0011888.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;

PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;

XX WPI: 1997-424969/39.

DR N-PSDB; AAW83814.

XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used  
PT to isolate antimicrobial compounds, and in vaccines against S.  
PT aureus infection

XX Claim 6; Page 310; 989pp; English.

XX The present sequence represents a Staphylococcus aureus protein of  
CC unknown function. The DNA sequence was isolated from a library of  
CC clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can  
CC be used in the construction of ribozymes and antisense sequences to  
CC control the expression of Staphylococcal genes. The DNA sequence is  
CC also useful as a source of regulatory elements for the control of  
CC bacterial gene expression. The present protein may be used to produce  
CC vaccines to enable a host to produce specific antibodies with  
CC antibacterial action. These vaccines and antibodies would protect

CC a host against invasion by *S. aureus*, and conditions relating to  
 CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled  
 CC skin syndrome, and toxic shock syndrome.

XX Sequence 115 AA;

alignment\_scores:  
 Quality: 34.00 Length: 8  
 Ratio: 4.857 Gaps: 0  
 Percent Similarity: 87.500 Percent Identity: 87.500

alignment\_block:  
 US-09-730-559B-45 x AAW27851 ..

Align seg 1/1 to: AAW27851 from: 1 to: 115

2 GCCTTAATATTATTCATAGATCGA 25  
 ||||| ||||| ||||| |||||  
 12 GlyLeuTyrLeuPheIleAspArg 19

seq\_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW11794

seq\_documentation\_block:  
 ID AAW11794 standard; Protein; 109 AA.

XX AC AAW11794;

XX DT 24-APR-1997 (first entry)

XX DE Early onset Alzheimer's disease gene product (uterine cancer).

XX KW Early onset Alzheimer's disease; EOAD; diagnosis; therapy;

XX KW expressed sequence tag; EST; uterus cancer.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Misc-difference 86 /note= "amino acid residue at position 86 is  
 FT unidentified owing to degeneracy of the  
 FT nucleotide sequence"

FT Misc-difference 89 /note= "amino acid residue at position 89 is  
 FT unidentified owing to degeneracy of the  
 FT nucleotide sequence"

FT Misc-difference 97 /note= "amino acid residue at position 97 is  
 FT unidentified owing to degeneracy of the  
 FT nucleotide sequence"

FT Misc-difference 105 /note= "amino acid residue at position 105 is  
 FT unidentified owing to degeneracy of the  
 FT nucleotide sequence"

XX WO9701573-A2.

XX PD 16-JAN-1997.

XX PF 26-JUN-1996; 96WO-US11002.

XX PR 28-JUN-1995; 95US-0000590.

XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.

XX PI Karran EH;

XX DR WPI; 1997-100161/09.

XX DR N-PSDB; AAT59474.

XX PT New Early Onset Alzheimer's Disease gene - used to develop prods.  
 PT for diagnosis, detection of pre-disposition to, or treatment of

PT Alzheimer's disease

PS Claim 3; Page 75; 97pp; English.

XX CC A set of polypeptides (AAW11792-97) have sequences deduced from the  
 CC 3 reading frames, of both strands, of a human uterine cancer cDNA  
 CC clone (AAT59474) that corresponds to an expressed sequence tag (EST)  
 CC of the early onset Alzheimer's disease (EOAD) gene. EOAD gene  
 CC products (see also AAW11768-91) can be expressed in host cells and  
 CC used to screen for agonists or antagonists useful in EOAD therapy,  
 CC or to raise antibodies useful in the diagnosis of EOAD or  
 CC predisposition to EOAD.

CC (N.B. in-frame stop codons in the 6 reading frames of the cDNA  
 CC clone are ignored in the translated polypeptide sequences given in  
 CC the specification)

XX SQ Sequence 109 AA;

alignment\_scores:  
 Quality: 33.00 Length: 7  
 Ratio: 4.714 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:  
 US-09-730-559B-45 x AAW11794 ..

Align seg 1/1 to: AAW11794 from: 1 to: 109

4 CTTAATATTATTCATAGATCG 24  
 ||||| ||||| ||||| |||||  
 34 LeuAsnIleLeuHisArgSer 40

seq\_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW11769

seq\_documentation\_block:  
 ID AAW11769 standard; Protein; 127 AA.

XX AC AAW11769;

XX DT 24-APR-1997 (first entry)

XX DE Early onset Alzheimer's disease gene product (amygdala).

XX KW Early onset Alzheimer's disease; EOAD; diagnosis; therapy;  
 XX KW expressed sequence tag; EST; amygdala.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Misc-difference 3 /note= "amino acid residue at position 3 is  
 FT unidentified owing to degeneracy of the  
 FT nucleotide sequence"

FT Misc-difference 46 /note= "amino acid residue at position 46 is  
 FT unidentified owing to degeneracy of the  
 FT nucleotide sequence"

FT Misc-difference 103 /note= "amino acid residue at position 103 is  
 FT unidentified owing to degeneracy of the  
 FT nucleotide sequence"

XX WO9701573-A2.

XX PD 16-JAN-1997.

XX PF 26-JUN-1996; 96WO-US11002.

XX PR 28-JUN-1995; 95US-0000590.

XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.



XX Karan EH;  
 XX WPI: 1997-100161/09.  
 DR N-PSDB; AAT59470.  
 XX  
 XX New Early Onset Alzheimer's Disease gene - used to develop prods.  
 PT for diagnosis, detection of pre-disposition to, or treatment of  
 PT Alzheimer's disease  
 XX  
 XX Claim 3; Page 55; 97pp; English.  
 XX  
 CC A set of polypeptides (AAW11768-73) have sequences deduced from the  
 CC 3 reading frames, of both strands, of a human amygdala cDNA clone  
 CC (AAT59470) that corresponds to an expressed sequence tag (EST) of the  
 CC early onset Alzheimer's disease (EOAD) gene. EOAD gene products  
 CC (see also AAW11774-97) can be expressed in host cells and used to  
 CC screen for agonists or antagonists useful in EOAD therapy, or to  
 CC raise antibodies useful in the diagnosis of EOAD or predisposition  
 CC to EOAD.  
 CC (N.B. in-frame stop codons in the 6 reading frames of the cDNA  
 CC clone are ignored in the translated polypeptide sequences given in  
 CC the specification)  
 XX  
 SQ Sequence 127 AA;  
  
 alignment\_scores:  
 Quality: 33.00 Length: 7  
 Ratio: 4.714 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 85.714  
  
 alignment\_block:  
 US-09-730-559B-45 x AAW11769 ..  
 Align seg 1/1 to: AAW11769 from: 1 to: 127  
  
 4 CTTAATATTATTATCATGATCG-24  
 |||||||:|||||||  
 25 LeuAsnIleLeuHisArgSer 31  
  
 seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW39625  
  
 seq\_documentation\_block:  
 ID AAW39625 standard; Protein: 390 AA.  
 XX  
 AC AAW39625;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 XX Human polypeptide SEQ ID NO 2770.  
 DE  
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 XX 26-JUL-2001.  
 PD  
 XX 26-DEC-2000; 2000WO-US34263.  
 PF  
 XX 21-JAN-2000; 2000US-0488725.  
 PR  
 PR 25-APR-2000; 2000US-0522317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI: 2001-442253/47.  
 DR N-PSDB; AAI58781.  
 XX  
 XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 PT  
 XX Example 4; SEQ ID NO 2770; 10078pp; English.  
 PS  
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAW42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemia and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SQ Sequence 390 AA;  
  
 alignment\_scores:  
 Quality: 33.00 Length: 6  
 Ratio: 5.500 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
  
 alignment\_block:  
 US-09-730-559B-45/rev x AAM39625 ..  
 Align seg 1/1 to: AAM39625 from: 1 to: 390  
  
 18 ATGATATATTATTAAAGCCC 1  
 |||||||:|||||||  
 35 MetAsnAsnIleLysPro 40  
  
 seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW41411  
  
 seq\_documentation\_block:  
 ID AAW41411 standard; Protein: 396 AA.  
 XX  
 AC AAW41411;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 XX Human polypeptide SEQ ID NO 6342.  
 DE  
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX

PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI: 2001-442253/47.  
DR N-PSDB; AA160567.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
XX Example 2; SEQ ID NO 6342; 10078pp; English.  
PS  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
XX Sequence 396 AA;  
SQ

alignment\_scores:  
Quality: 33.00 Length: 6  
Ratio: 5.500 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-730-559B-45/rev x AAM41411 ..  
Align seg 1/1 to: AAM41411 from: 1 to: 396  
18 ATGAATATATATAGCCC 1  
41 MetAsnAllylPro 46

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: AAG67128

seq\_documentation\_block:  
ID AAG67128 standard; Protein; 404 AA.  
XX  
AC AAG67128;  
XX  
XX 13-NOV-2001 (first entry)  
XX  
XX Amino acid sequence of a human enzyme.  
XX  
XX Human; enzyme; cancer; neurological disorder; epilepsy; stroke;

KW Alzheimer's disease; Pick's disease; Huntington's disease; dementia;  
KW multiple sclerosis; Parkinson's disease; amyotrophic lateral sclerosis;  
KW meningitis; schizophrenic disorder; neuroskeletal disorder; allergy;  
KW Addison's disease; autoimmune disease; anemia; asthma; Crohn's disease;  
KW adult respiratory distress syndrome; atopic dermatitis; psoriasis;  
KW diabetes mellitus; osteoporosis; pancreatitis; rheumatoid arthritis;  
KW infection; genetic disorder; muscular dystrophy; Gaucher's disease;  
KW Huntington's chorea; sickle cell anemia; thalassemia; atherosclerosis;  
KW Von Willebrand's disease; Wilms' tumour; cell proliferative disorder;  
KW leukemia; hepatitis; cirrhosis; arteriosclerosis; gene therapy.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Modified-site 136  
FT /note= "potential phosphorylation site"  
FT Modified-site 155  
FT /note= "potential phosphorylation site"  
FT Modified-site 192  
FT /note= "potential phosphorylation site"  
FT Modified-site 293  
FT /note= "potential phosphorylation site"  
FT Modified-site 301  
FT /note= "potential phosphorylation site"  
XX  
XX WO200164896-A2.  
XX  
XX 07-SEP-2001.  
XX  
XX 01-MAR-2001; 2001WO-US06806.  
XX  
XX 01-MAR-2000; 2000US-0186307.  
PR 28-MAR-2000; 2000US-0192532.  
PR 30-MAR-2000; 2000US-0193578.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Tang YT, Lu DAM, Bandman O, Yue H, Azimzai Y, Lal P, Burford N;  
PI Baughn MR;  
XX  
XX WPI: 2001-550184/61.  
DR N-PSDB; AAH75156.  
XX  
XX Novel human enzyme molecule useful for treating and preventing, e.g.,  
PT cancer, genetic disorders, neurological disorders, autoimmune and  
PT inflammatory disorders -  
XX  
XX Claim 1; Page 118; 154pp; English.  
PS  
XX The present sequence represents a human enzyme. The enzyme polynucleotide  
CC and polypeptide are useful for diagnosis, treatment and prevention of  
CC cancers, neurological disorders (e.g. epilepsy, stroke, Alzheimer's  
CC disease, Pick's disease, Huntington's disease, dementia, multiple  
CC sclerosis, Parkinson's disease, amyotrophic lateral sclerosis, bacterial  
CC and viral meningitis, schizophrenic disorders and neuroskeletal  
CC disorders), autoimmune/inflammatory disorders (e.g. allergies,  
CC Addison's disease, autoimmune diseases, adult respiratory distress  
CC syndrome, anemia, asthma, Crohn's disease, atopic dermatitis, diabetes  
CC mellitus, osteoporosis, pancreatitis, psoriasis, rheumatoid arthritis,  
CC and viral, bacterial, fungal, parasitic, protozoal and helminthic  
CC infections), genetic disorder (e.g. Duchenne and Becker muscular  
CC dystrophy, Gaucher's disease, Huntington's chorea, sickle cell anemia,  
CC thalassemia, Von Willebrand's disease and Wilms' tumour), and cell  
CC proliferative disorder (e.g. atherosclerosis, leukemia, hepatitis,  
CC cirrhosis, and arteriosclerosis). The polynucleotide is also useful in  
CC somatic or germline gene therapy.  
XX  
XX Sequence 404 AA;  
SQ

alignment\_scores:  
Quality: 33.00 Length: 6  
Ratio: 5.500 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-730-559b-45/rev x AAG67128 ..

Align seg 1/1 to: AAG67128 from: 1 to: 404

18 ATGAATAATATTAAAGCCC 1

|||||

1 MetAsnAsnIleLysPro 6

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: AAB93837

## seq\_documentation\_block:

ID AAB93837 standard; Protein: 434 AA.

XX AAB93837;

DT 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:13667.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX Claim 8; SEQ ID 13667; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX Sequence 434 AA;

## alignment\_scores:

Quality: 33.00 Length: 6

Ratio: 5.500 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-730-559b-45/rev x AAB93837 ..

Align seg 1/1 to: AAB93837 from: 1 to: 434

18 ATGAATAATATTAAAGCCC 1

|||||

42 MetAsnAsnIleLysPro 47

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: AAW16678

## seq\_documentation\_block:

ID AAW16678 standard; Protein: 548 AA.

XX AAW16678;

DT 20-AUG-1997 (first entry)

XX Lawsonia intracellularis GroEL.

KW Intestinal disease; porcine proliferative enteropathy; vaccine;

KW GroEL; immunotherapy; antibody; diagnosis; heat shock protein.

XX Lawsonia intracellularis.

OS WO9720050-A1.

PN 05-JUN-1997.

XX 29-NOV-1996; 96WO-AU00767.

XX 30-NOV-1995; 95AU-0006911.

PR 30-NOV-1995; 95AU-0006910.

XX (DARA-) DARATECH PTY LTD.

PA (PIGR-) PIG RES & DEV CORP.

XX Hasse D, Panaccio M;

PI WPI; 1997-310605/28.

DR N-PSDB; AAT69201.

XX Vaccine for treating or preventing Lawsonia intracellularis  
PT infection - especially in pigs, containing non-pathogenic form of  
PT bacterium or its components

XX Claim 10; Page 38-42; 94pp; English.

CC The GroEL heat shock protein (AAW16678) of Lawsonia intracellularis  
CC can be used in vaccines to protect birds and animals against  
CC intestinal diseases, esp. to protect pigs against porcine  
CC proliferative enteropathy. It is the expression product of a DNA  
CC molecule (AAH69201) obtd. by screening an L. intracellularis library  
CC with rabbit anti-L. intracellularis sera. GroEL, GroES (AAW16679)  
CC and other L. intracellularis polypeptides (AAW16680-85) can be  
CC administered as recombinant polypeptides or expressed as  
CC recombinant vaccines utilising bacterial, fungal or viral vectors.  
CC Antibodies raised against the polypeptides may be useful in  
CC immunotherapy, diagnosis of infection and detection.

XX Sequence 548 AA;

alignment\_scores:  
Quality: 33.00 Length: 8  
Ratio: 4.125 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.500

alignment\_block:  
US-09-730-559B-45 x AAU16678 ..

Align seg 1/1 to: AAU16678 from: 1 to: 548

1 GGGCTTAATATATTATTCATAGATCG 24  
|||||  
439 GlyLeuGluIleLeuArgSer 446

seq\_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAU10933

seq\_documentation\_block:

ID AAB10933 standard; protein; 599 AA.

AC AAB10933;

DT 07-FEB-2001 (first entry)

DE S. pombe rsn3p protein fragment.

KW RNA polymerase I transcription factor TIF-1A; rsn3p; antitumor; cancer;  
KW antiproliferative; cell proliferation; treatment; tissue regeneration.

OS Schizosaccharomyces pombe.

PN WO200055316-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-DE00767.

PR 17-MAR-1999; 99DE-1011992.

PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

PI Grummt I, Vingron M;

DR WPI; 2000-587527/55.

XX New DNA encoding the transcription factor TIF-1A, useful for preventing  
XX or treating diseases associated with abnormal cell proliferation,  
XX particularly tumors

PS Example 1; Fig 1; 38pp; German.

XX This invention describes a novel DNA sequence (I) that encodes the RNA  
XX polymerase I transcription factor TIF-1A which has antitumor,  
XX antiproliferative and proliferation-inducing activity. The invention also  
XX describes (1) DNA (Ia) encoding a protein (II) with the biological  
XX activity of TIF-1A; (2) a ribozyme (R) corresponding to (I) or (Ia) and  
XX able to bind specifically to, and cleave, its transcribed RNA so as to  
XX reduce or inhibit synthesis of the corresponding protein; (3) an  
XX antisense RNA (AS) with binding properties similar to R; (4) an  
XX expression vector that contains (I), (Ia) or sequences that encode R or  
XX AS; (5) host cells containing the vectors of (4); (6) TIF-1A or (II)  
XX cells of (6); (7) preparation of TIF-1A or (II) by culturing  
XX cells of (6); (8) ligands that bind to TIF-1A or (II); (9) antagonists  
XX that weaken or block the activity of TIF-1A or (II); (10) a diagnostic  
XX method for detecting abnormal TIF-1A expression; and (11) kit for  
XX carrying out the method in (10). (I), and similar sequences that encode  
XX proteins with equivalent activity, expression vectors containing them, or  
XX their expression products are used to treat or prevent disorders  
XX associated with reduced cellular proliferation, to stimulate cellular  
XX proliferation, and to promote tissue regeneration, e.g. after injury or  
XX radiation therapy. Ribozymes, antisense sequences directed against (I),  
XX also ligands and antagonists of TIF-1A, are used to treat or prevent  
XX disorders associated with excessive cellular proliferation and to inhibit  
XX proliferation, especially in treatment of cancers. (I) and specific

CC ligands for TIF-1A (particularly antibodies (Ab)) are also useful for  
CC diagnosis of altered TIF-1A expression by (in)direct determination of the  
CC concentration, length and/or sequence of TIF-1A or its mRNA, e.g. for  
CC detecting mutations. Ab can also be used for immunoprecipitation of  
CC TIF-1A and for isolation of related sequences from cDNA expression  
CC libraries. (I) allows recombinant production of TIF-1A in sufficient  
CC quantities for therapeutic use.

SQ Sequence 599 AA;

alignment\_scores:  
Quality: 33.00 Length: 8  
Ratio: 4.125 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:

US-09-730-559B-45 x AAB10933 ..

Align seg 1/1 to: AAB10933 from: 1 to: 599

1 GGGCTTAATATATTATTCATAGATCG 24

|||||  
488 GlyLeuGluIleLeuHisArgSer 495

seq\_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAU51340

seq\_documentation\_block:

ID AAU51340 standard; Protein; 634 AA.

AC AAU51340;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #12236.

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

DR N-PSDB; AAS59551.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
XX vaccinating against and diagnosing infections, especially useful for  
XX treating acne vulgaris

PS Example 1; SEQ ID No 12535; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
XX polypeptides. The proteins and their associated DNA sequences are used in  
XX the treatment, prevention and diagnosis of medical conditions caused by  
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
XX P. acnes is also involved in infections of bone, joints and the central

CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of *P. acnes* in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for *P. acnes* proteins. These antibodies can be used to  
CC downregulate expression and activity of *P. acnes* polypeptides and  
CC therefore treat *P. acnes* infections. The antibodies may also be used as  
CC diagnostic agents for determining *P. acnes* presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 634 AA;

alignment\_scores:                   Quality: 33.00                   Length: 7  
                                      Ratio: 4.714                   Gaps: 0  
Percent Similarity: 100.000       Percent Identity: 85.714

alignment\_block:

US-09-730-559B-45 x AAU51340   ..

Align seg 1/1 to: AAU51340 from: 1 to: 634

2 GGCTTATATATTCATAGAT 22  
|||||:::|||||  
58 GlyLeuValLeuPheIleAsp 64

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:AAU58611

seq\_documentation\_block:

ID\_AAR58611 standard; Protein; 1045 AA.

XX AC AAR58611;

XX DT 28-APR-1995 (first entry)

XX DE Yeast HMG-CoA reductase 2.

XX KW HMG-CoA reductase 2; 3-hydroxy-3-methyl:glutaryl coenzyme A;  
XX squalene; sterol.

XX OS Saccharomyces cerevisiae.

XX PN US5349126-A.

XX PD 20-SEP-1994.

XX PF 12-OCT-1990; 90US-0596467.

XX PR 12-OCT-1990; 90US-0596467.

XX PR 14-AUG-1992; 92US-0934374.

XX PA (STAD ) AMOCO CORP.

XX PI Chappell J, Saunders CA, Wolf FR;

XX DR WPI; 1994-302280/37.

XX DR N-PSDB; AAQ70611.

XX PT Transgenic plants with increased insect resistance - by  
XX transformation with DNA encoding HMG-CoA activity to increase  
XX squalene and sterol accumulation

XX PS Disclosure; Page 53; 58pp; English.

XX CC AAQ70611 codes for a yeast HMG-CoA reductase 2, AAR58611. Using the  
XX A. tumefaciens vector pKYLX71, AAQ70611 was used to produce transgenic  
XX plants with increased HMG-CoA reductase activity. Plants with

CC increased HMG-CoA reductase activity were found to have greater  
CC squalene and sterol accumulation, and therefore increased insect  
CC resistance.

XX Sequence 1045 AA;

alignment\_scores:                   Quality: 33.00                   Length: 7  
                                      Ratio: 4.714                   Gaps: 0  
Percent Similarity: 100.000       Percent Identity: 85.714

alignment\_block:

US-09-730-559B-45 x AAR58611   ..

Align seg 1/1 to: AAR58611 from: 1 to: 1045

4 CTTAATATATTCATAGATCG 24  
:::|||||  
360 IleAsnIleIleHisArgSer 366

---

OM of: US-09-730-559B-45 to: Issued Patents\_AA:\* out\_format : pfs  
Date: Sep 8, 2002 8:11 AM  
About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters: -MODEL=frame+np2.model -DEV=xlh  
-DB=cgn2\_1/USPTO\_spool/US09730559/runat\_03092002\_150444\_3263/app\_query.fasta\_1.86  
-DB-Issued\_Patents\_AA -QFMT=fastan -SUFFIX=ra1 -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEX=7.000 -START=1 -MATRIX=blomsm62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09730559\_cgn1\_1.8 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-730-559B-45  
Query length: 26  
Database: Issued Patents\_AA:\*  
Database sequences: 231628  
Database length: 24425594  
Search time (sec): 29.910000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-596-467-6 +		33.00	107.92	15.21	1045
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-934-374-6 +		33.00	107.92	15.21	1045
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-783-861C-6 +		33.00	107.92	15.21	1045
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-09-433-994-2 +		32.00	109.58	23.81	540
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-596-467-4 +		32.00	103.81	25.54	1054
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-934-374-4 +		32.00	103.81	25.54	1054
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-783-861C-2 +		32.00	103.81	25.54	1054
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-471-780C-86 -		31.00	129.90	29.66	32
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-467-282B-86 -		31.00	129.90	29.66	32
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-471-282A-86 -		31.00	129.90	29.66	32
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-466-710C-86 -		31.00	129.90	29.66	32
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-468-739C-86 -		31.00	129.90	29.66	32
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-3 -		31.00	104.49	40.45	610
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-5 -		31.00	104.49	40.45	610
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-11 -		31.00	104.27	40.56	626
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-12 -		31.00	100.29	42.58	993
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-449-645A-17 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-702-367A-17 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-2 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-17 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-19 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-20 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-21 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-22 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-23 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-24 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-25 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-26 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-27 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-28 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-29 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-30 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-31 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-32 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-33 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-34 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-35 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-36 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-37 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-38 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-39 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-40 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-41 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-42 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-43 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-44 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-45 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-46 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-47 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-48 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-49 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-50 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-51 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-52 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-53 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-54 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-55 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-56 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-57 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-58 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-59 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-60 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-61 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-62 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-63 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-64 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-65 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-66 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-67 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-68 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-69 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-70 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-71 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-72 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-73 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-74 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-75 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-76 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-77 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-78 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-79 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-80 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-81 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-82 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-83 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-84 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-85 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-86 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-87 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-88 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-89 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-90 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-91 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-92 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-93 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-94 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-95 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-96 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-97 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-98 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-99 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-100 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-101 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-102 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-103 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-104 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-105 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-106 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-107 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-108 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-109 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-110 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-111 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-112 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-113 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-114 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-115 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-116 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-117 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-118 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-119 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-120 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-121 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-122 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-123 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-124 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-125 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-126 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-127 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-128 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-129 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-130 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-131 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-132 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-133 -		31.00	100.25	42.61	998
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:					

```
; TITLE OF INVENTION: Process and Composition for Increasing
; TITLE OF INVENTION: Sterol Accumulation in Higher Plants
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Goldsmith, Shore, Suker & Milnamow
; STREET: 180 N. Stetson St.
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/934,374
; FILING DATE: 19920814
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Gamson, Edward P.
; REGISTRATION NUMBER: 29,381
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 616-5400
; TELEFAX: 312 616-5460
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1045 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-934-374-6

alignment_scores:
  Quality: 33.00      Length: 7
  Ratio: 4.714      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 85.714

alignment_block:
US-09-730-559B-45 x US-07-934-374-6 ..
Align seg 1/1 to: US-07-934-374-6 from: 1 to: 1045

4 CTTAATATTATTCATAGATCG 24
:::|||||
360 IleAsnIleIleHisArgSer 366

seq_name: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:US-07-783-861C-6

seq_documentation_block:
; Sequence 6, Application US/07783861C
; Patent No. 5460949
; GENERAL INFORMATION:
; APPLICANT: Saunders, Court A.
; APPLICANT: Wolf, Fred R.
; APPLICANT: Mukharji, Indrani
; TITLE OF INVENTION: A Method and Composition for Increasing
; TITLE OF INVENTION: the Accumulation of Squalene and Specific Sterols in
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corp., Patents and Licensing Dept.
; STREET: 200 East Randolph St.
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60680-0703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/783,861C
; FILING DATE: 19911028
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/613,380
; FILING DATE: 15-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, No. 5460949vall B.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 856-7180
; TELEFAX: 312 856-4972
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1045 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-783-861C-6

alignment_scores:
  Quality: 33.00      Length: 7
  Ratio: 4.714      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 85.714

alignment_block:
US-09-730-559B-45 x US-07-783-861C-6 ..
Align seg 1/1 to: US-07-783-861C-6 from: 1 to: 1045

4 CTTAATATTATTCATAGATCG 24
:::|||||
360 IleAsnIleIleHisArgSer 366

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:US-09-433-994-2

seq_documentation_block:
; Sequence 2, Application US/09433994
; Patent No. 6326172
; GENERAL INFORMATION:
; APPLICANT: Zalacain, Magdalena
; APPLICANT: Biswas, Sanjoy
; APPLICANT: Warren, Patrick V.
; APPLICANT: Ingraham, Karen I.
; APPLICANT: Chalker, Alison F.
; APPLICANT: So, Chi Young
; APPLICANT: Holmes, David J.
; APPLICANT: Van Horn, Stephanie
; TITLE OF INVENTION: ytgp
; FILE REFERENCE: GM10246
; CURRENT APPLICATION NUMBER: US/09/433,994
; CURRENT FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-09-433-994-2

alignment_scores:
  Quality: 32.00      Length: 8
  Ratio: 4.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 62.500

alignment_block:
US-09-730-559B-45/rev x US-09-433-994-2 ..
Align seg 1/1 to: US-09-433-994-2 from: 1 to: 540

24 CGATCTATGAAATATTAAGCCC 1
```





```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/613,380
; FILING DATE: 15-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, No. 5460949vall B.
; TELEPHONE: 312 856-7180
; TELEFAX: 312 856-4972
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1054 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-783-861C-2

alignment_scores:
  Quality: 32.00      Length: 7
  Ratio: 4.571       Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 71.429

alignment_block:
US-09-730-559B-45 x US-07-783-861C-2 ..

Align seg 1/1 to: US-07-783-861C-2 from: 1 to: 1054

4 CTTAATATTATTATGATCG 24
  ::::::::::::::::::::
361 MetAsnValIleHIsArgSer 367

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-471-780C-86

seq_documentation_block:
; Sequence 86, Application US/08471780C
; Patent No. 5759808
; GENERAL INFORMATION:
; APPLICANT: Casterman, Cecile
; APPLICANT: Hamers, Raymond
; TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,780C
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,944
; FILING DATE: 17-AUG-1993
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: FR 93401310.3
; FILING DATE: 21-MAY-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93401310.3
; FILING DATE: 21-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 04958.0008-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400

; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-467-282B-86

; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-467-282B-86

; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-471-780C-86

alignment_scores:
  Quality: 31.00      Length: 6
  Ratio: 5.167       Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 83.333

alignment_block:
US-09-730-559B-45/rev x US-08-471-780C-86 ..

Align seg 1/1 to: US-08-471-780C-86 from: 1 to: 32

18 ATGAATAATATTAAAGCCC 1
17 MetAsnAsnLeuLysPro 22

seq_name: /cgn2_5/ptodata/1/iaa/5A_COMB.pep:US-08-467-282B-86

seq_documentation_block:
; Sequence 86, Application US/08467282B
; Patent No. 5800988
; GENERAL INFORMATION:
; APPLICANT: Casterman, Cecile
; APPLICANT: Hamers, Raymond
; TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,282B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,944
; FILING DATE: 17-AUG-1993
; APPLICATION NUMBER: FR 92402326.0
; FILING DATE: 21-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93401310.3
; FILING DATE: 21-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 04958.0008-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-467-282B-86
```

alignment\_scores:  
Quality: 31.00 Length: 6  
Ratio: 5.167 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 83.333

alignment\_block:  
US-09-730-559B-45/rev x US-08-467-282B-86 ..

Align seg 1/1 to: US-08-467-282B-86 from: 1 to: 32

18 ATGAATAATATTAAAGCCC 1  
|||||:|||||  
17 MetAsnAsnLeuLysPro 22

seq\_name: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pap:US-08-471-282A-86

seq\_documentation\_block:  
; Sequence 86, Application US/08471282A  
; Patent No. 5840853  
; GENERAL INFORMATION:  
; APPLICANT: Casterman, Cecile  
; APPLICANT: Hamers, Raymond  
; TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains  
; NUMBER OF SEQUENCES: 130  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471.282A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/106.944  
; FILING DATE: 17-AUG-1993  
; APPLICATION NUMBER: FR 92402326.0  
; FILING DATE: 21-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 93401310.3  
; FILING DATE: 21-MAY-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Potter, Jane E.R.  
; REGISTRATION NUMBER: 33,332  
; REFERENCE/DOCKET NUMBER: 04958.0008-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-408-4400  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 86:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-471-282A-86

alignment\_scores:  
Quality: 31.00 Length: 6  
Ratio: 5.167 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 83.333

alignment\_block:

US-09-730-559B-45/rev x US-08-471-282A-86 ..

Align seg 1/1 to: US-08-471-282A-86 from: 1 to: 32

18 ATGAATAATATTAAAGCCC 1  
|||||:|||||  
17 MetAsnAsnLeuLysPro 22

seq\_name: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pap:US-08-466-710C-86

seq\_documentation\_block:  
; Sequence 86, Application US/08466710C  
; Patent No. 5874541  
; GENERAL INFORMATION:  
; APPLICANT: Casterman, Cecile  
; APPLICANT: Hamers, Raymond  
; TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains  
; NUMBER OF SEQUENCES: 130  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,710C  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/106,944  
; FILING DATE: 17-AUG-1993  
; APPLICATION NUMBER: FR 92402326.0  
; FILING DATE: 21-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 93401310.3  
; FILING DATE: 21-MAY-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Potter, Jane E.R.  
; REGISTRATION NUMBER: 33,332  
; REFERENCE/DOCKET NUMBER: 04958.0008-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 86:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-466-710C-86

alignment\_scores:  
Quality: 31.00 Length: 6  
Ratio: 5.167 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 83.333

alignment\_block:

US-09-730-559B-45/rev x US-08-466-710C-86 ..

Align seg 1/1 to: US-08-466-710C-86 from: 1 to: 32

18 ATGAATAATATTAAAGCCC 1  
|||||:|||||  
17 MetAsnAsnLeuLysPro 22

```
seq_name: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:US-08-468-739C-86
seq_documentation_block:
; Sequence 86, Application US/08468739C
; Patent No. 6015695
; GENERAL INFORMATION:
; APPLICANT: Casterman, Cecile
; APPLICANT: Hamers, Raymond
; TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,739C
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,944
; FILING DATE: 17-AUG-1993
; APPLICATION NUMBER: FR 92402326.0
; FILING DATE: 21-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93401310.3
; FILING DATE: 21-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 04958.0008-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-468-739C-86

alignment_scores:
    Quality: 31.00      Length: 6
    Ratio: 5.167        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 83.333

alignment_block:
US-09-730-559B-45/rev x US-08-468-739C-86      ..
Align seg 1/1 to: US-08-468-739C-86 from: 1 to: 32

18 ATGAATAATATTAAAGCCC 1
|||||:|||||:|||||
17 MetAsnAsnLeuLysPro 22

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:US-08-368-776A-3
seq_documentation_block:
; Sequence 3, Application US/08368776A
; Patent No. 6300482
; GENERAL INFORMATION:
; APPLICANT: Ciossek, Thomas
; APPLICANT: Ullrich, Axel
```

```
; APPLICANT: Millauer, Birgit
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS
; AND TREATMENT OF MDK1
; TITLE OF INVENTION: SIGNAL TRANSDUCTION
; DISORDERS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/368,776A
; FILING DATE: January 3, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 208/007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-368-776A-3

alignment_scores:
    Quality: 31.00      Length: 7
    Ratio: 4.429        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 71.429

alignment_block:
US-09-730-559B-45/rev x US-08-368-776A-3      ..
Align seg 1/1 to: US-08-368-776A-3 from: 1 to: 610

21 TCTATGAATAATATTAAAGCCC 1
|||||:|||||:|||||
501 SerIleAsnAsnLeuLysPro 507

seq_name: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:PCT-US96-00419-3
seq_documentation_block:
; Sequence 3, Application PC/TUS9600419
; GENERAL INFORMATION:
; APPLICANT: Thomas Ciossek, Axel Ullrich, Birgit
; APPLICANT: Millauer
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT OF
; MDK1 SIGNAL TRANSDUCTION DISORDERS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
```

```
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: IBM P.C. DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00419
; FILING DATE: January 3, 1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 208/007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US96-00419-3
;
; alignment_scores:
; Quality: 31.00 Length: 7
; Ratio: 4.429 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 71.429
;
; alignment_block:
; US-09-730-559B-45/rev x PCT-US96-00419-3 ..
;
; Align seg 1/1 to: PCT-US96-00419-3 from: 1 to: 610
;
; 21 TCTATGAATAATATTAGCCC 1
; 501 SerIleAsnLeuLysPro 507
;
; seq_name: /cgn2_5/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-5
;
; seq_documentation_block:
; Sequence 5, Application US/08368776A
; Patent No. 630482
; GENERAL INFORMATION:
; APPLICANT: Ciossek, Thomas
; APPLICANT: Ullrich, Axel
; APPLICANT: Millauer, Birgit
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF MDK1
; TITLE OF INVENTION: SIGNAL TRANSDUCTION
; TITLE OF INVENTION: DISORDERS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/368,776A
; FILING DATE: January 3, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 208/007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 626 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-368-776A-5
;
; alignment_scores:
; Quality: 31.00 Length: 7
; Ratio: 4.429 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 71.429
;
; alignment_block:
; US-09-730-559B-45/rev x US-08-368-776A-5 ..
;
; Align seg 1/1 to: US-08-368-776A-5 from: 1 to: 626
;
; 21 TCTATGAATAATATTAGCCC 1
; 501 SerIleAsnLeuLysPro 507
```

---

Sequence	Strnd	Orig	ZScore	Escore	Len	Documentation
p1r2:S44834	-	38.00	137.13	4.41	343	F54H12.3 protein - Caenorhabditis
p1r2:T27230	-	38.00	137.13	4.41	343	hypothetical protein Y57G11C.18
p1r2:T32093	-	38.00	136.94	4.42	351	hypothetical protein K09F6.1 -
p1r1:AA0811	+	36.00	130.14	12.32	301	myosin-light-chain kinase (EC 2
p1r2:T21558	+	35.00	111.38	23.04	1785	hypothetical protein F2966.3b
p1r2:T21559	+	35.00	110.74	23.14	1929	hypothetical protein F2966.3a
p1r2:B70335	+	34.00	126.24	33.61	182	conserved hypothetical protein
p1r2:B70335	+	34.00	123.35	34.33	258	hypothetical protein [imported]
p1r2:H90051	+	34.00	123.35	34.33	258	hypothetical protein [imported]
p1r2:F64238	-	34.00	119.64	35.29	404	hypothetical protein MG349 - My
p1r2:B87350	+	34.00	117.24	35.92	540	hypothetical protein CC0813 [im
p1r2:A11088	+	33.00	120.23	57.21	231	transcription regulators, Fnr/C
p1r2:T37924	+	33.00	112.35	60.64	599	rna polymerase I specific trans
p1r2:B30239	+	33.00	107.74	62.74	1045	hydroxymethylglutaryl-CoA redu
p1r2:G82595	+	32.00	112.43	98.67	364	ABC transporter phosphate bindi
p1r2:C87134	+	32.00	112.21	98.84	374	conserved hypothetical protein
p1r2:T33186	-	32.00	111.21	99.57	422	hypothetical protein F22F7.2 -
p1r1:S28051	+	32.00	110.88	99.81	439	transcription factor GCN5 - yea
p1r2:F98720	-	32.00	110.87	99.82	440	protein H25P06.2b [imported] -
p1r2:T32123	+	32.00	110.57	100.47	456	hypothetical protein H25P06.2a
p1r2:T32123	+	32.00	109.99	100.47	489	hypothetical protein H25P06.2b
p1r2:T23124	+	32.00	109.94	100.95	529	polysaccharide biosynthesis pro
p1r2:A93178	-	32.00	109.17	101.08	540	polysaccharide transporter [imp
p1r2:F98044	+	32.00	108.31	101.72	599	hypothetical protein FXJ13.80 -
p1r2:T51548	+	32.00	108.03	101.94	620	ATPase-like protein [imported]
p1r2:H85431	+	32.00	108.03	101.94	620	ATPase-like ser/thr protein kinase
p1r2:T50148	+	32.00	107.78	102.13	639	probable ser/thr protein kinase
p1r2:A84563	+	32.00	106.99	102.72	703	probable AAA-type ATPase [impo
p1r2:T19974	+	32.00	105.81	103.62	811	hypothetical protein C46F11.4 -
p1r2:AC3070	+	32.00	105.19	104.10	874	ATP-dependent Clp proteinase, A
p1r2:F98216	+	32.00	105.06	104.19	887	endeopeptidase clp ATP-binding
p1r2:D81821	+	32.00	104.50	104.63	950	glycine dehydrogenase (decarbox
p1r2:A30239	+	32.00	103.64	105.30	1054	hydroxymethylglutaryl-CoA redu
p1r2:S46754	+	32.00	102.37	106.28	1228	hypothetical protein YH1555 -
p1r2:E90098	+	32.00	101.97	106.60	1289	RNA polymerase III largest sub
p1r2:T27552	-	32.00	96.97	116.62	2361	hypothetical protein F45E4.4 -
p1r2:S33599	-	31.00	118.15	154.03	112	Ig gamma chain - Arabian camel
p1r2:T12721	+	31.00	117.11	155.21	127	hypothetical protein 5 - Methan
p1r2:B26956	+	31.00	115.36	157.24	163	ribosomal protein S7 - Micrococ
p1r2:AH3408	+	31.00	115.05	157.60	163	molybdopterin (mpt) converting
p1r2:H64643	-	31.00	112.99	160.51	209	hypothetical protein HJ0991 - H
p1r2:H64300	-	31.00	112.57	160.51	220	hypothetical protein MJ0008 - M

```
24 CGATCTATGTAATAATATTAAAGCCC 1
      :::::::::::::::::::::
222 LysSerMetAsnValLysPro 229
      seq_name: pir2:T32093

seq_documentation_block:
hypothetical protein K09F6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000
C:Accession: T32093
R:Scheet, P.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid K09F6.
A:Reference number: Z21122
A:Accession: T32093
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-351 <SCH>
A:Cross-references: EMBL:AF016683; PIDN:AAB66192.1; GSPDB:GN00020; CESP:K09F6.1
A:Experimental source: strain Bristol N2; clone K09F6
C:Genetics:
A:Gene: CESP:K09F6.1
A:Map position: 2
C:Superfamily: Caenorhabditis elegans hypothetical protein K09F6.1

alignment_scores:
      Quality: 38.00      Length: 8
      Ratio: 4.750      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 75.000

alignment_block:
US-09-730-559B-45/rev x T32093 ..
Align seg 1/1 to: T32093 from: 1 to: 351
24 CGATCTATGTAATAATATTAAAGCCC 1
      :::::::::::::::::::::
263 LysSerMetAsnValLysPro 270
      seq_name: pir1:A40811

seq_documentation_block:
myosin-light-chain kinase (EC 2.7.1.117) A - slime mold (Dictyostelium discoideum)
N:Alternate names: MLCK-A
C:Species: Dictyostelium discoideum
C:Date: 10-Apr-1992 #sequence_revision 21-Jan-1997 #text_change 11-Jun-1999
C:Accession: A40811; A37125
R:Tan, J.L.; Spudich, J.A.
J. Biol. Chem. 266, 16044-16049, 1991
A:Title: Characterization and bacterial expression of the Dictyostelium myosin light chain
A:Reference number: A40811; MUID:91340753
A:Accession: A40811
A:Molecule type: mRNA
A:Residues: 1-301 <TAN>
A:Cross-references: GB:M64176; NID:g1498249; PIDN:AAB06337.1; PID:g1498250
R:Tan, J.L.; Spudich, J.A.
J. Biol. Chem. 265, 13818-13824, 1990
A:Title: Dictyostelium myosin light chain kinase. Purification and characterization.
A:Reference number: A37125; MUID:90337997
A:Accession: A37125
A:Molecule type: protein
A:Residues: 9-12, 1', 14-19; 163-167, 'S', 169-179; 192-198 <TA2>
C:Genetics:
A:Gene: mlkA
C:Function:
A:Description: catalyzes the formation of peptidyl-serine-phosphate or peptidyl-threonin
d in itself
A:Pathway: cytokinesis; fruiting body formation
A:Note: not activated by Ca2+/calmodulin in contrast with MLCK from higher eukaryotes
C:Superfamily: kinase-related transforming protein; protein kinase homology

C:Keywords: ATP; autophosphorylation; cell division; phosphoprotein; phosphotransfera
F:6-265/Domain: protein kinase homology <KIN>
F:14-22/Region: protein kinase ATP-binding motif
F:267-293/Domain: inhibitory #status predicted <INH>
F:37/Active site: Lys #status predicted
F:296/Binding site: phosphate (Thr) (covalent) (by autophosphorylation) #status predi
```



A:Experimental source: clone F29G6  
C:Genetics:  
A:Gene: CESP:F29G6.3a  
A:Map position: X  
A:Introns: 10/1; 43/3; 61/3; 88/3; 110/3; 144/2; 183/3; 218/3; 252/3; 288/3; 1083/3; 121

alignment\_scores:  
Quality: 35.00 Length: 8  
Ratio: 4.375 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 62.500

alignment\_block:

US-09-730-559B-45 x T21559 ..

Align seg 1/1 to: T21559 from: 1 to: 1929

1 GGGCTTAATATTATTCATAGATCG 24  
|||||:|||||:|||||:|||||:  
261 GlyValAsnValLeuHisArgSer 268

seq\_name: pir2:B70335

seq\_documentation\_block:

conserved hypothetical protein aq\_389 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999

C:Accession: B70335

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oviatt, J.; ...

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666

A:Accession: B70335

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-182 <AQF>

A:Cross-references: GB:AE000688; NID:g2983063; PIDN:AAC06687.1; PID:g2983080; GB:AE00065

A:Experimental source: strain VF5

C:Genetics:

A:Gene: aq\_389

alignment\_scores:

Quality: 34.00 Length: 7  
Ratio: 4.857 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:

US-09-730-559B-45 x B70335 ..

Align seg 1/1 to: B70335 from: 1 to: 182

1 GGGCTTAATATTATTCATAGA 21  
|||||:|||||:|||||:|||||:  
43 GlyLeuAsnIleIleHisLys 49

seq\_name: pir2:H90051

seq\_documentation\_block:

hypothetical protein [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C:Accession: H90051

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; ...

C:Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: H90051

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-258 <KUR>  
A:Cross-references: GB:BA000018; PID:g13702437; PIDN:BA043578.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA2275  
C:Superfamily: Staphylococcus aureus hypothetical protein lpl1

alignment\_scores:

Quality: 34.00 Length: 8  
Ratio: 4.250 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 62.500

alignment\_block:

US-09-730-559B-45 x H90051 ..

Align seg 1/1 to: H90051 from: 1 to: 258

2 GGCTTAATATTATTCATAGCGA 25  
|||||:|||||:|||||:|||||:  
94 GlyLeuValLeuTyValAspArg 101

seq\_name: pir2:F64238

seq\_documentation\_block:

hypothetical protein MG349 - Mycoplasma genitalium

C:Species: Mycoplasma genitalium

C>Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 07-Dec-1999

C:Accession: F64238

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, C.A.; Venter, J.C.

Science 270, 397-403, 1995

A:Title: The minimal gene complement of Mycoplasma genitalium.

A:Reference number: A64200; MUID:96026346

A:Accession: F64238

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-404 <TIGR>

A:Cross-references: GB:U39719; GB:L43967; NID:g1046055; PID:g1046056; TIGR:MG349

A:Experimental source: strain G-37

C:Genetics:

A:Genetic code: SGC3

alignment\_scores:

Quality: 34.00 Length: 7  
Ratio: 4.857 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:

US-09-730-559B-45/rev x F64238 ..

Align seg 1/1 to: F64238 from: 1 to: 404

21 TCTATGAATTAATTAAGCCC 1  
|||||:|||||:|||||:|||||:  
383 SerLeuAsnIleLysPro 389

seq\_name: pir2:B87350

seq\_documentation\_block:

hypothetical protein CC0813 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: B87350

R:Nierman, M.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

n, J.; Emolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete genome sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: B87350

<p>A;Status: preliminary</p> <p>A;Molecule type: DNA</p> <p>A;Residues: 1-540 &lt;STO&gt;</p> <p>A;Cross-references: GB:AE005673; NID:g13422062; PIDN:AAK22798.1; GSPDB:GN00148</p> <p>C;Genetics:</p> <p>A;Gene: CC0813</p>	<p>alignment_scores:</p> <p>Quality: 34.00 Length: 8</p> <p>Ratio: 4.857 Gaps: 0</p> <p>Percent Similarity: 87.500 Percent Identity: 87.500</p> <p>alignment_block:</p> <p>US-09-730-559B-45 x H87350 ..</p> <p>Align seg 1/1 to: B87350 from: 1 to: 540</p> <p>1 GGCTTAATATTATTCATAGATCGA 25</p> <p>     :     :     </p> <p>431 GlyLeuIleuPheTyrAspArg 438</p> <p>seq_name: pir2:A11088</p> <p>seq_documentation_block:</p> <p>transcription regulators, Fnr/Crp family homolog lmo0112 [imported] - Listeria monocytogenes</p> <p>C;Species: Listeria monocytogenes</p> <p>C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001</p> <p>C;Accession: A11088</p> <p>R;Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, H.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.</p> <p>Science 294, 849-852, 2001</p> <p>A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.</p> <p>A;Reference number: AB1077; MUID:21537279; PMID:11679669</p> <p>A;Accession: A11088</p> <p>A;Status: preliminary</p> <p>A;Molecule type: DNA</p> <p>A;Residues: 1-231 &lt;GLA&gt;</p> <p>A;Cross-references: GB:NC_003210; PIDN:CAC98327.1; PID:g16409471; GSPDB:GN00177</p> <p>A;Experimental source: strain EGD-e</p> <p>C;Genetics:</p> <p>A;Gene: lmo0112</p>	<p>alignment_scores:</p> <p>Quality: 33.00 Length: 8</p> <p>Ratio: 4.714 Gaps: 0</p> <p>Percent Similarity: 87.500 Percent Identity: 75.000</p> <p>alignment_block:</p> <p>US-09-730-559B-45 x A11088 ..</p> <p>Align seg 1/1 to: A11088 from: 1 to: 231</p> <p>2 GGCTTAATATTATTCATAGATCGA 25</p> <p>    :     :     </p> <p>96 GlyThrValLeuPheIleAspArg 103</p> <p>seq_name: pir2:T37924</p> <p>seq_documentation_block:</p> <p>rna polymerase I specific transcription initiati on factor - fission yeast (Schizosaccharomyces pombe)</p> <p>C;Species: Schizosaccharomyces pombe</p> <p>C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999</p> <p>C;Accession: T37924</p> <p>R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.</p> <p>submitted to the EMBL Data Library, December 1995</p> <p>A;Reference number: Z21754</p> <p>A;Accession: T37924</p> <p>A;Status: preliminary</p> <p>A;Gene: translated from GB/EMBL/DBJ</p>	<p>A;Molecule type: DNA</p> <p>A;Residues: 1-599 &lt;CON&gt;</p> <p>A;Cross-references: EMBL:Z68198; PIDN:CAA92389.1; GSPDB:GN00066; SPDB:SPAC1866.11c</p> <p>A;Experimental source: strain 972h-; cosmid c1866</p> <p>C;Genetics:</p> <p>A;Gene: SPDB:SPAC1866.11c</p> <p>A;Map position: 1</p>	<p>alignment_scores:</p> <p>Quality: 33.00 Length: 8</p> <p>Ratio: 4.125 Gaps: 0</p> <p>Percent Similarity: 100.000 Percent Identity: 75.000</p> <p>alignment_block:</p> <p>US-09-730-559B-45 x T37924 ..</p> <p>Align seg 1/1 to: T37924 from: 1 to: 599</p> <p>1 GGGCTTAATATTATTCATAGATCG 24</p> <p>     :     :     </p> <p>488 GlyLeuGluIleuHisArgSer 495</p> <p>seq_name: pir2:B30239</p> <p>seq_documentation_block:</p> <p>hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) II - yeast (Saccharomyces cerevisiae)</p> <p>N;Alternate names: protein L9324.2; protein YLR450w</p> <p>C;Species: Saccharomyces cerevisiae</p> <p>C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 05-May-2000</p> <p>C;Accession: B30239; B24317; S55972</p> <p>R;Basson, M.E.; Thorsness, M.; Finer-Moore, J.; Stroud, R.M.; Rine, J.</p> <p>Mol. Cell. Biol. 8, 3797-3808, 1988</p> <p>A;Title: Structural and functional conservation between yeast and human 3-hydroxy-3-methylglutaryl-CoA reductase</p> <p>A;Reference number: A93105; MUID:89127221</p> <p>A;Accession: B30239</p> <p>A;Molecule type: DNA</p> <p>A;Residues: 1-1045 &lt;BAS&gt;</p> <p>A;Cross-references: EMBL:M22255; NID:g171687; PIDN:AAA34677.1; PID:g171688</p> <p>R;Basson, M.E.; Thorsness, M.; Rine, J.</p> <p>Proc. Natl. Acad. Sci. U.S.A. 83, 5563-5567, 1986</p> <p>A;Title: Saccharomyces cerevisiae contains two functional genes encoding 3-hydroxy-3-methylglutaryl-CoA reductase</p> <p>A;Reference number: A94112; MUID:86287298</p> <p>A;Accession: B24317</p> <p>A;Molecule type: DNA</p> <p>A;Residues: 772-961 &lt;BAS2&gt;</p> <p>R;Du, Z.</p> <p>submitted to the EMBL Data Library, March 1995</p> <p>A;Description: The sequence of S. cerevisiae cosmid 9324.</p> <p>A;Reference number: S55966</p> <p>A;Accession: S55972</p> <p>A;Molecule type: DNA</p> <p>A;Residues: 1-1045 &lt;DUZ&gt;</p> <p>A;Cross-references: GB:U22382; NID:g717059; PIDN:AAB67527.1; PID:g717061; MIPS:YLR450</p> <p>C;Genetics:</p> <p>A;Gene: SGD:HMG2</p> <p>A;Cross-references: SGD:S0004442; MIPS:YLR450w</p> <p>A;Map position: 12R</p> <p>C;Superfamily: hydroxymethylglutaryl-CoA reductase I</p> <p>F;29-45/Domain: transmembrane #status predicted &lt;TM1&gt;</p> <p>F;248-264/Domain: transmembrane #status predicted &lt;TM2&gt;</p> <p>F;331-347/Domain: transmembrane #status predicted &lt;TM3&gt;</p> <p>F;402-418/Domain: transmembrane #status predicted &lt;TM4&gt;</p> <p>F;502-518/Domain: transmembrane #status predicted &lt;TM5&gt;</p> <p>F;681-697/Domain: transmembrane #status predicted &lt;TM6&gt;</p> <p>F;991-1007/Domain: transmembrane #status predicted &lt;TM7&gt;</p> <p>alignment_scores:</p> <p>Quality: 33.00 Length: 7</p> <p>Ratio: 4.714 Gaps: 0</p> <p>Percent Similarity: 100.000 Percent Identity: 85.714</p>
--	--	---	--	---

## alignment\_block:

US-09-730-559B-45 x B30239 ..

Align seg 1/1 to: B30239 from: 1 to: 1045

4 CTTAATATTATTCATAGATCG 24

:::|||||

360 IleAsnIleIleHisArgSer 366

seq\_name: pir2:G82595

## seq\_documentation\_block:

ABC transporter phosphate binding protein XF2141 [imported] - Xylella fastidiosa (strain

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

C:Accession: G82595

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: G82595

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-364 &lt;SIM&gt;

A:Cross-references: GB:AE004028; GB:AE003849; NID:g9107266; PIDN:AAF84940.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H

as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2141

C:Superfamily: phosphate-repressible phosphate-binding protein

## alignment\_scores:

Quality: 32.00 Length: 7

Ratio: 4.571 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 71.429

## alignment\_block:

US-09-730-559B-45 x G82595 ..

Align seg 1/1 to: G82595 from: 1 to: 364

4 CTTAATATTATTCATAGATCG 24

:::|||||

179 IleAsnIleValHisArgSer 185

seq\_name: pir2:C87134

## seq\_documentation\_block:

Conserved hypothetical protein ML1802 [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: C87134

R:Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho

eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq

A:Title: Massive gene decay in the leprosy bacillus.

A:Reference number: A86909; MUID:21128732; PMID:11234002

A:Accession: C87134

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-374 &lt;STO&gt;

A:Cross-references: GB:AL450380; NID:g13093513; PIDN:CAC30755.1; GSPDB:GN00147

C:Genetics:

A:Gene: ML1802

## alignment\_scores:

Quality: 32.00 Length: 7

Ratio: 4.571 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 85.714

## alignment\_block:

US-09-730-559B-45 x C87134 ..

Align seg 1/1 to: C87134 from: 1 to: 374

5 TTAATATTATTCATAGATCGA 25

|||||

54 LeuValIeuPheIleAspArg 60

\_\_\_\_\_

OM of: US-09-730-559b-45 to: SwissProt\_40:\* out\_format : pfs  
 Date: Sep 8, 2002 8:13 AM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
 -MODEL=framed1.n2p.model -DEV=xlh  
 -Q/cgn2\_1/USPTO.spool/US09730559/runat\_03092002.150446.3360/app\_query.fasta\_1.86  
 -DB=SwissProt\_40 -QPMT=fastan -SUFFIX=rsf -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000  
 -GAPOP=6.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
 -TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
 -NORM=US09730559.CCNI\_1\_13 -NCPU=6 -ICPU=3 -LONGLOG  
 -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-730-559b-45  
 Query length: 26  
 Database: SwissProt\_40:\*  
 Database sequences: 105224  
 Database length: 38719550  
 Search time (sec): 28.210000

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
SwissProt_40:YMD4_CAEEL	38.00	136.67	1.89	343	P34458 caenorhabditis elegans
SwissProt_40:KMLC_DICDI	36.00	129.84	5.26	295	P25323 dictyostelium discoideum
SwissProt_40:Y349_MYCGE	34.00	119.40	14.66	404	P47591 mycoplasma genitalium
SwissProt_40:Y4Q2_SCHPO	33.00	112.34	24.46	599	Q10110 schizosaccharomyces cerevisiae
SwissProt_40:HMD2_YEAST	33.00	108.01	24.44	1045	P12684 saccharomyces cerevisiae
SwissProt_40:OXDA_MOUSE	32.00	112.61	40.89	346	P18894 mus musculus (mouse)
SwissProt_40:PSTS_XYLP	32.00	112.22	40.88	364	Q9PBK3 xylella fastidiosa
SwissProt_40:GCG5_YEAST	32.00	110.76	40.87	439	Q03330 saccharomyces cerevisiae
SwissProt_40:HMD1_YEAST	32.00	103.94	40.80	1054	P12683 saccharomyces cerevisiae
SwissProt_40:YHVS_YEAST	32.00	102.75	40.79	1228	P38851 saccharomyces cerevisiae
SwissProt_40:RS7_MCLU	31.00	114.81	68.37	156	P09898 micrococcus luteus (m
SwissProt_40:MOAE_RHIL	31.00	114.47	68.37	163	Q8YGA6 brucella melitensis
SwissProt_40:RL15_QUESE	31.00	112.84	68.34	201	O82712 quercus suber (cork
SwissProt_40:Y008_METJA	31.00	112.14	68.33	220	Q60319 methanococcus jannasch
SwissProt_40:VG07_BPP22	31.00	111.83	68.32	229	Q01074 bacteriophage p22. dn
SwissProt_40:FAEL_ECOLI	31.00	110.93	68.31	257	P33784 escherichia coli. k88
SwissProt_40:YQ23_CAEEL	31.00	108.20	68.27	365	Q09452 caenorhabditis elegans
SwissProt_40:Y4Q2_RHISN	31.00	106.60	68.24	448	P55628 rhizobium sp. (strain
SwissProt_40:IMDH_PYRAB	31.00	105.99	68.23	485	Q9UY49 pyrococcus abyssal. in
SwissProt_40:IMDH_PYRAB	31.00	105.99	68.23	485	P42851 pyrococcus furiosus
SwissProt_40:IMDH_PYRHO	31.00	105.97	68.23	486	O58045 pyrococcus horikoshii
SwissProt_40:IMDH_METJA	31.00	105.81	68.23	496	Q59011 methanococcus jannasch
SwissProt_40:AMPA_RICPR	31.00	105.75	68.23	500	P27888 rickettsia prowazekii
SwissProt_40:GUAA_ECOLI	31.00	105.37	68.22	525	P04079 escherichia coli. gnd
SwissProt_40:YMX1_CAEEL	31.00	105.05	68.22	547	P34509 caenorhabditis elegans
SwissProt_40:FRP1_SCHPO	31.00	104.81	68.21	564	Q04800 schizosaccharomyces p
SwissProt_40:NODQ_AZOBR	31.00	104.07	68.20	620	P28604 a nodq bifunctional e
SwissProt_40:SMX2_YEAST	31.00	102.19	68.17	790	P32909 saccharomyces cerevis
SwissProt_40:CLP_TRYBB	31.00	101.46	68.16	868	P31543 trypanosoma brucei br
SwissProt_40:CLPB_TREPA	31.00	101.37	68.16	878	O83110 treponema pallidum. d
SwissProt_40:EPAT_CHICK	31.00	100.41	68.14	993	O42422 gallus gallus (chicken
SwissProt_40:EPAT_HUMAN	31.00	100.37	68.14	998	Q15375 homo sapiens (human)
SwissProt_40:EPAT_MOUSE	31.00	100.37	68.14	998	Q61772 mus musculus (mouse)
SwissProt_40:EPAT_RAT	31.00	100.37	68.14	998	Q61772 mus musculus (mouse)
SwissProt_40:CALF_HUMAN	31.00	96.68	68.08	1603	P54759 rattus norvegicus (rat)
SwissProt_40:YLU9_CAEEL	30.00	112.29	114.21	129	P34403 caenorhabditis elegans
SwissProt_40:RS7_MYCLE	30.00	110.87	114.17	155	P30764 mycobacterium leprae.
SwissProt_40:RS7_MYCXM	30.00	110.87	114.17	155	P41193 mycobacterium smegmati
SwissProt_40:RS7_MYCTO	30.00	110.87	114.17	155	P41194 mycobacterium tubercul
SwissProt_40:RS7_MYCBO	30.00	110.82	114.17	156	Q53539 mycobacterium bovis. 3

SwissProt_40:RS7_STRECO	30.00	110.82	114.17	156	! Q910k4 streptomyces coelic
SwissProt_40:MOAE_RHIL	30.00	110.67	114.17	159	! Q984p0 rhizobium loti (me
SwissProt_40:VP11_RDVA	30.00	109.66	114.14	181	! P28890 rice dwarf virus (i
SwissProt_40:VP11_RDVF	30.00	109.32	114.13	189	! Q85442 rice dwarf virus (i
SwissProt_40:Y278_ARATH	30.00	107.63	114.09	235	! O82246 arabidopsis thaliana

seq\_name: SwissProt\_40:YMD4\_CAEEL

seq\_documentation\_block:  
 ID YMD4\_CAEEL STANDARD; PRT; 343 AA.  
 AC P34458;

DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 40.8 kDa protein F54H12.4 in chromosome III.  
 GN F54H12.4.

OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
 RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden K.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K.,  
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 RA Wodhams P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

CC EMBL; L25599; AAA28053.1; -  
 CC PIR; S44834; S44834.  
 DR WormPep: F54H12.4; CE00546.  
 DR Hypothetical protein.  
 KW Hypothetical protein.  
 SQ SEQUENCE 343 AA; 40799 MW; F636DCB35BA96F96 CRC64;

alignment\_scores:  
 Quality: 38.00 Length: 8  
 Ratio: 4.750 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:  
 US-09-730-559B-45/rev x YMD4\_CAEEL ..  
 Align seg 1/1 to: YMD4\_CAEEL from: 1 to: 343

24 CGATCTGATGATATATTAAGCCC 1  
 ::::|||||:|||||:|||||:|||||  
 222 LysSerMetAsnValLysPro 229

seq\_name: SwissProt\_40:KMLC\_DICDI  
 seq\_documentation\_block:  
 ID KMLC\_DICDI STANDARD; PRT; 295 AA.  
 AC P25323;

DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE MYOSIN light chain kinase (EC 2.7.1.117) (MLCK).  
GN MLCK.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX3;  
RX MEDLINE=91340753; PubMed=1651931;  
RA Tan J.L., Spudich J.A.;  
RT "Characterization and bacterial expression of the Dictyostelium  
myosin light chain kinase cDNA. Identification of an autoinhibitory  
domain.";  
RL J. Biol. Chem. 266:16044-16049(1991).  
RN [2]  
RP REVISIONS.  
RA Spudich J.A.;  
RN Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP PARTIAL SEQUENCE.  
RC STRAIN=AX3;  
RX MEDLINE=90337997; PubMed=2380188;  
RA Tan J.L., Spudich J.A.;  
RT "Dictyostelium myosin light chain kinase. Purification and  
characterization.";  
RL J. Biol. Chem. 265:13818-13824(1990).  
CC -1- FUNCTION: PHOSPHORYLATES A SPECIFIC SERINE IN THE N-TERMINUS OF A  
MYOSIN LIGHT CHAIN.  
CC -1- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin  
light-chain] phosphate.  
CC -1- ENZYME REGULATION: POSSESSES AN AUTOINHIBITORY DOMAIN.  
CC AUTOPHOSPHORYLATION APPEARS TO INCREASE THE ENZYMIC ACTIVITY.  
CC DOES NOT HAVE A CALMODULIN-BINDING DOMAIN.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC CAMK SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; M64176; AAB06337.1; -  
DR PIR; A40811; A40811.  
DR PIR; A37125; A37125.  
DR HSP; Q63450; IAO6.  
DR DictyDb; DD01034; mlka.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR SMART; SM00220; S\_TRC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
KW Phosphorylation.  
FT DOMAIN 8 265 PROTEIN KINASE.  
FT DOMAIN 264 295 AUTOINHIBITORY DOMAIN.  
FT NP\_BIND 14 22 ATP (BY SIMILARITY).  
FT BINDING 37 37 ATP (BY SIMILARITY).  
FT ACT\_SITE 130 130 BY SIMILARITY.  
SQ SEQUENCE 295 AA; 33406 MW; 546CAEED8F6ECD0B CRC64;

alignment\_scores:  
Quality: 36.00 Length: 7  
Ratio: 5.143 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:  
US-09-730-559B-45 x KMLC\_DICDI ..  
Align seg 1/1 to: KMLC\_DICDI from: 1 to: 295  
1 GGGCTTAATATATTATTCATAGA 21  
|||||  
123 GlyLeuAsnIleValHisArg 129  
seq\_name: SwissProt\_40:Y349\_MYCGE  
seq\_documentation\_block:  
ID Y349\_MYCGE STANDARD; PRT; 404 AA.  
AC P47591;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein MG349.  
GN MG349.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=96026346; PubMed=7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RT "The minimal gene complement of Mycoplasma genitalium.";  
RL Science 270:397-403(1995).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; U39716; AAC71574.1; -  
DR TIGR; MG349; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 404 AA; 47774 MW; 37023D6A8F92CAE4 CRC64;

alignment\_scores:  
Quality: 34.00 Length: 7  
Ratio: 4.857 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:  
US-09-730-559B-45/rev x Y349\_MYCGE ..  
Align seg 1/1 to: Y349\_MYCGE from: 1 to: 404  
21 TCTATGAATAATTAAGCC 1  
|||||  
383 SerLeuAsnIleValHisPro 389  
seq\_name: SwissProt\_40:YAOB\_SCHPO  
seq\_documentation\_block:  
ID YAOB\_SCHPO STANDARD; PRT; 599 AA.  
AC Q10110;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical 68.6 kDa protein C1866.11c in chromosome I.  
GN SPAC1866.11C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
ON NCBI\_TaxID=4896;  
RX NCBI  
RN SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
CC -|- SIMILARITY: TO YEAST RRN3 AND C.ELEGANS C36E8.1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z68198; CAA92389.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 599 AA; 68554 MW; 8351B6612CE768EF CRC64;  
-----  
alignment\_scores:  
Quality: 33.00 Length: 8  
Ratio: 4.125 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.000  
-----  
alignment\_block:  
US-09-730-559B-45 x YAOB\_SCHPO ..  
Align seg 1/1 to: YAOB\_SCHPO from: 1 to: 599  
1 GGGCTTAATATATTCATAGATCG 24  
|||||:|||||:|||||  
488 GlyLeuGlutLeuHisArgSer 495  
seq\_name: SwissProt\_40:HMD2\_YEAST  
seq\_documentation\_block:  
ID HMD2\_YEAST STANDARD; PRT; 1045 AA.  
AC P12684;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE 3-hydroxy-3-methylglutaryl-coenzyme A reductase 2 (EC 1.1.1.34) (HMG-  
DE CoA reductase 2).  
GN HMG2 OR YLR450W OR L9324.2.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
ON NCBI\_TaxID=4932;  
RX NCBI  
RN SEQUENCE FROM N.A.  
RC MEDLINE=89127221; PubMed=3065625;  
RA Basson M.E., Thorness M., Finer-Moore J., Stroud R.M., Rine J.;  
RT "Structural and functional conservation between yeast and human 3-  
RT hydroxy-3-methylglutaryl coenzyme A reductases, the rate-limiting  
RT enzyme of sterol biosynthesis".  
RL Mol. Cell. Biol. 8:3797-3808(1988).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,  
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,  
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,  
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,  
RA Rifkin L., Riles L., Taich A., Trevasakis E., Vignati D.,  
-----  
alignment\_scores:  
Quality: 33.00 Length: 7  
Ratio: 4.714 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714  
-----

```

alignment_block:
US-09-730-559B-45 x HMD2_YEAST  ..

Align seg 1/1 to: HMD2_YEAST from: 1 to: 1045

      4 CTTAATATTATTATTCATAGTCG 24
      :::|||||
360 IleaSnIlelleHlsArgSer 366

seq_name: SwissProt_40:OXDA_MOUSE

seq_documentation_block:
ID   OXDA_MOUSE STANDARD;          PRT;   346 AA.
AC   P18994; Q64465;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   D-amino acid oxidase (EC 1.4.3.3) (DAMOX) (DAO) (DAAO).
GN   DAO OR DAO1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RC   SEQUENCE FROM N.A.
RP   TISSUE=Kidney;
RX   MEDLINE=90382679; PubMed=1976103;
RA   Tada M., Fukui K., Momoi K., Miyake Y.;
RT   "Cloning and expression of a cDNA encoding mouse kidney D-amino acid
RT   oxidase."
RL   Gene 90:293-297(1990).
[2]
RP   SEQUENCE FROM N.A.
RA   Sasaki M.;
RL   Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
[3]
RX   MUTAGENESIS OF GLY-182.
RX   MEDLINE=92385496; PubMed=1355365;
RA   Sasaki M., Konno R., Nishio M., Niwa A., Yasumura Y., Enami J.;
RT   "A single-base-pair substitution abolishes D-amino-acid oxidase
RT   activity in the mouse."
RL   Biochim. Biophys. Acta 1139:315-318(1992).
CC   -1- FUNCTION: COULD ACT AS A DETOXIFYING AGENT WHICH REMOVES D-AMINO
CC   ACIDS ACCUMULATED DURING AGING. ACTS ON A VARIETY OF D-AMINO ACIDS
CC   WITH A PREFERENCE FOR THOSE HAVING SMALL HYDROPHOBIC SIDE CHAINS
CC   FOLLOWED BY THOSE BEARING POLAR, AROMATIC, AND BASIC GROUPS. DOES
CC   NOT ACT ON ACIDIC AMINO ACIDS.
CC   -1- CATALYTIC ACTIVITY: A D-amino acid + H(2)O + O(2) = a 2-oxo acid +
CC   NH(3) + H(2)O(2).
CC   -1- COFACTOR: FAD.
CC   -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -1- SUBCELLULAR LOCATION: Peroxisomal.
CC   -1- SIMILARITY: BELONGS TO THE DAMOX/DASOX FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC   EMBL; M32299; AAA39367.1; -
CC   EMBL; D10210; BAA01062.1; -
CC   EMBL; D10211; BAA01063.1; -
CC   PIR; JH0185; JH0185.
CC   HSP; P00371; IAN9.
CC   MGD; MGI:94859; Dao1.
CC   InterPro; IPR000927; DAO.
CC   Pfam; PF01266; DAO; 1.
CC   PROSITE; PS00342; MICROBODIES_CTER; 1.
CC   PROSITE; PS00677; DAO; 1.
CC   Oxidoreductase; Flavoprotein; FAD; Peroxisome.

```



```

CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE004028; AAF84940.1; -.
DR HSSP; P06128; 11XH.
DR InterPro; IPR002370; PstS.
DR Pfam; PF01449; PstS; 1.
KW Phosphate transport; Transport; Periplasmic; Signal;
KW Complete proteome.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 364 PHOSPHATE-BINDING PERIPLASMIC PROTEIN.
SQ SEQUENCE 364 AA; 38736 MW; 8609CFAA159D4277 CRC64;

alignment_scores:
  Quality: 32.00 Length: 7
  Ratio: 4.571 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 71.429

alignment_block:
US-09-730-559B-45 x PSTS_XYLFA ..
Align seg 1/1 to: PSTS_XYLFA from: 1 to: 364
4 CTTAATATTATTCATGATCG 24
:::|||||:::|||||
179 IleaSnlleValHISArgSer 185

seq_name: SwissProt_40:GCN5_YEAST

seq_documentation_block:
ID GCN5_YEAST STANDARD; PRT; 439 AA.
AC Q03330;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Transcriptional activator GCN5.
GN GCN5 OR ADA4 OR YGR252W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93011009; PubMed-1396595;
RA Georgakopoulos T., Thireos G.;
RT "Two distinct yeast transcriptional activators require the function
RT of the GCN5 protein to promote normal levels of transcription.";
RL EMBO J. 11:4145-4152(1992).
RN [2]
RP SEQUENCE OF 1-170 FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE-97279234; PubMed-9133742;
RA Peroli F., Carignani G., Pavanello A., Guerreiro P., Azevedo D.,
RA Rodrigues-Pousada C., Melchiorretto P., Panzeri L.,
RA Agostoni Carbone M.L.;
RT "Analysis of a 17.9 kb region from Saccharomyces cerevisiae
RT chromosome VII reveals the presence of eight open reading frames,
RT including BRFL1 (TFIIB70) and GCN5 genes.";
RL Yeast 13:373-377(1997).
RN [3]
RP SEQUENCE OF 170-439 FROM N.A.
RC STRAIN=S288C;
RX MEDLINE-97279233; PubMed-9133741;
RA Mazzoni C., Ruzzi M., Rinaldi T., Solinas F., Montebove F.,

```

```

RA Frontali L.;
RT "Sequence analysis of a 10.5 kb DNA fragment from the yeast
RT chromosome VII reveals the presence of three new open reading frames
RT and of a tRNAthr gene.";
RL Yeast 13:369-372(1997).
RN [4]
RP ASSOCIATION WITH ADA2.
RX MEDLINE-95045371; PubMed-7957049;
RA Marcus G.A., Silverman N., Berger S.L., Horiuchi J., Guarente L.;
RT "Functional similarity and physical association between GCN5 and
RT ADA2: putative transcriptional adaptors.";
RL EMBO J. 13:4807-4815(1994).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 99-262.
RX MEDLINE-99362688; PubMed=10430873;
RA Trievel R.C., Rojas J.R., Sterner D.E., Venkataramani R.N., Wang L.,
RA Zhou J., Allis C.D., Berger S.L., Marmorstein R.;
RT "Crystal structure and mechanism of histone acetylation of the yeast
RT GCN5 transcriptional coactivator.";
Proc. Natl. Acad. Sci. U.S.A. 96:8931-8936(1999).
CC -!- FUNCTION: GENERAL TRANSCRIPTIONAL ACTIVATOR OPERATING IN CONCERT
CC WITH CERTAIN OTHER DNA-BINDING TRANSCRIPTIONAL ACTIVATORS SUCH AS
CC GCN4 OR HAP2/3/4. FUNCTION AS AN HISTONE ACETYLTRANSFERASE (HAT)
CC TO PROMOTE TRANSCRIPTIONAL ACTIVATION. HAS A STRONG PREFERENCE FOR
CC LYSINE 14 OF H3 AND A SOMEWHAT LOWER PREFERENCE FOR LYSINES 8 AND
CC 16 OF HISTONE H4.
CC -!- SUBUNIT: HETERODIMER WITH ADA2. PART OF THE ADA/GCN5 COMPLEX THAT
CC CONSISTS OF HFI1/ADAL, ADA2, ADA3, SPI20/ADA5 AND GCN5.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68628; CAA48602.1; -.
DR EMBL; X73037; CAA97281.1; -.
DR EMBL; X99228; CAA67614.1; -.
DR PIR; S28051; S28051.
DR PDB; 1VGH; 02-AUG-99.
DR TRANSFAC; T02145; -.
DR SGD; S0003484; GCN5.
DR InterPro; IPR000182; Acetyltransf_GCN5.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00583; Acetyltransf; 1.
DR Pfam; PF00439; bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS0014; BROMODOMAIN_2; 1.
KW Transcription regulation; DNA-binding; Activator; Trans-acting factor;
KW 3D-structure.
FT ACT_SITE 173 173 GENERAL BASE.
FT DOMAIN 344 414 BROMODOMAIN.
SQ SEQUENCE 439 AA; 51069 MW; 3200730DDC7EF70D CRC64;

alignment_scores:
  Quality: 32.00 Length: 8
  Ratio: 4.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 62.500

alignment_block:
US-09-730-559B-45/rev x GCN5_YEAST ..
Align seg 1/1 to: GCN5_YEAST from: 1 to: 439
24 CGATCTATGAATAATATTAAAGCCC 1
:::|||||:|||||

```

```
295 LysAspLeuAsnAsnIleIysPro 302
seq_name: SwissProt_40:HMD1_YEAST
seq_documentation_block:
ID HMD1_YEAST STANDARD; PRT: 1054 AA.
AC P12683;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1 (EC 1.1.1.34) (HMG-
DE CoA reductase 1).
GN HMG1 OR YML075C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89127221; PubMed=3065625;
RA Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J.;
RT "Structural and functional conservation between yeast and human 3-
RT hydroxy-3-methylglutaryl coenzyme A reductases, the rate-limiting
RT enzyme of sterol biosynthesis.";
RL Mol. Cell. Biol. 8:3797-3808(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Brown D., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (Oct-1994) to the EMBL/GenBank/DBSJ databases.
RN [3]
RP SEQUENCE OF 776-965 FROM N.A.
RX MEDLINE=86287298; PubMed=3526336;
RA Basson M.E., Thorsness M., Rine J.;
RT "Saccharomyces cerevisiae contains two functional genes encoding 3-
RT hydroxy-3-methylglutaryl-coenzyme A reductase.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5563-5567(1986).
CC -!- FUNCTION: THIS TRANSMEMBRANE GLYCOPROTEIN IS INVOLVED IN THE
CC ENZYME OF CHOLESTEROL BIOSYNTHESIS. IT IS THE RATE-LIMITING
CC ENZYME OF THE STEROL BIOSYNTHESIS.
CC -!- CATALYTIC ACTIVITY: (R)-mevalonate + CoA + 2 NADP(+) = (S)-3-
CC hydroxy-3-methylglutaryl-CoA + 2 NADPH.
CC -!- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
CC reticulum.
CC -!- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22002; AAA34676.1; -
DR EMBL; Z46373; CA86503.1; -
DR PIR; A24317; A24317.
DR PIR; A30239; A30239.
DR PIR; S48822; S48822.
DR SGD; S0004540; HMG1.
DR InterPro; IPR002202; HMG-CoA_red.
DR InterPro; IPR000731; HMGCR_patched_5TM.
DR Pfam; PF00368; HMG-CoA_red; 1.
DR PRINTS; PR00071; HMGCOARDTASE.
DR PROSITE; PS00066; HMG_COA_REDUCTASE_1; 1.
DR PROSITE; PS00318; HMG_COA_REDUCTASE_2; 1.
DR PROSITE; PS01192; HMG_COA_REDUCTASE_3; 1.
DR PROSITE; PS00065; HMG_COA_REDUCTASE_4; 1.
DR PROSITE; PS00156; SSD; 1.
KW Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;
KW Cholesterol biosynthesis; NADP; Multigene family.
FT DOMAIN 1 524 MEMBRANE-BOUND.
```

```
FT DOMAIN 525 617 LINKER.
FT DOMAIN 618 1054 CATALYTIC.
FT TRANSMEM 27 53 POTENTIAL.
FT TRANSMEM 187 211 POTENTIAL.
FT TRANSMEM 242 266 POTENTIAL.
FT TRANSMEM 300 324 POTENTIAL.
FT TRANSMEM 332 357 POTENTIAL.
FT TRANSMEM 398 422 POTENTIAL.
FT TRANSMEM 499 524 BY SIMILARITY.
FT ACT_SITE 714 714 BY SIMILARITY.
FT ACT_SITE 924 924 BY SIMILARITY.
FT ACT_SITE 1020 1020 GENERAL BASE (BY SIMILARITY).
FT CARBOHYD 115 115 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 452 452 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 490 490 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1054 AA; 115625 MW; 2B624944FB7B2DD0 CRC64;

alignment_scores:
Quality: 32.00 Length: 7
Ratio: 4.571 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 71.429

alignment_block:
US-09-730-559b-45 x HMD1_YEAST ..
Align seg 1/1 to: HMD1_YEAST from: 1 to: 1054

4 CTTATATTATTCATAGATCG 24
:::||||:|||||||
361 MetAsnValIleHisArgSer 367

seq_name: SwissProt_40:YHV5_YEAST
seq_documentation_block:
ID YHV5_YEAST STANDARD; PRT: 1228 AA.
AC P38851;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical 143.6 kDa protein in SPO16-REC104 intergenic region.
GN YHR155W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favella A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.";
RL Science 265:2077-2082(1994).
CC -!- SIMILARITY: STRONG, TO YEAST SIP3. SOME, TO S.POMBE SPAC19A8.02.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
CC -----
DR EMBL; U10397; AABG8977.1; -.
DR PIR; S46754; S46754.
DR SGD; S0001198; YHR155W.
DR InterPro; IPR001849; PH.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS0003; PH_DOMAIN; 1.
KW Hypothetical protein.
FT DOMAIN 308 421 PH.
SQ SEQUENCE 1228 AA; 143583 MW; C8872EAE8270A4B6 CRC64;

alignment_scores:
  Quality: 32.00 Length: 8
  Ratio: 4.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 75.000

alignment_block:
US-09-730-559B-45 x YHV5_YEAST ..
Align seg 1/1 to: YHV5_YEAST from: 1 to: 1228
1 GGGCTTAATATTATTCATAGTCG 24
|||||:|||||:|||||
1106 GlyLeuGlnIleMetHisArgSer 1113

seq_name: SwissProt_40:RS7_MICLU

seq_documentation_block:
ID RS7_MICLU STANDARD; PRT; 156 AA.
AC P09898;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE 30S ribosomal protein S7.
DE RPSG.
GN Micrococcus luteus (Micrococcus lysodeikticus).
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Micrococcaceae; Micrococcus.
OX NCBI_TaxID=1270;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007427; PubMed=3654584;
RA Ohama T., Yamao F., Muto A., Osawa S.;
RT "organization and codon usage of the streptomycin operon in
RT Micrococcus luteus, a bacterium with a high genomic G + C content.";
RL J. Bacteriol. 169:4770-4777(1987).
CC -1- FUNCTION: PROTEIN S7 BINDS SPECIFICALLY TO PART OF THE 3' END OF
CC 16S RIBOSOMAL RNA (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S7P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

|||||:|||||:|||||:|||||:
82 GlyLeuThrIleHisArg 88
seq_name: SwissProt_40:RL15_QUESU
seq_documentation_block:
ID RL15_QUESU STANDARD; PRT; 201 AA.
AC O82712;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L15.
GN RPL15.
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fagales; Fagaceae; Quercus.
OX NCBI_TaxID=58331;
RN [1]
SEQUENCE FROM N.A.
RP Huguet G., Pla M., Verdaguier D., Molinas M.;
RA Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RL - SIMILARITY: BELONGS TO THE L15E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ001346; CAA04690.1; -
DR InterPro; IPR000439; Ribosomal L15e.
DR Pfam; PF00827; Ribosomal L15e; 1.
DR PROSITE; PS01194; RIBOSOMAL_L15E; 1.
KW Ribosomal protein.
SQ SEQUENCE 201 AA; 23135 MW; F77F66B61886CF3D CRC64;

alignment_scores:
Quality: 31.00 Length: 7
Ratio: 4.429 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 71.429

alignment_block:
US-09-730-559B-45 x RL15_QUESU ..
Align seg 1/1 to: RL15_QUESU from: 1 to: 201

4 CTTAATATTATTCATAGTCG 24
|||||:|||||:|||||:|||||:
33 LeuAsnValIleHisArgAla 39
seq_name: SwissProt_40:Y008_METJA
seq_documentation_block:
ID Y008_METJA STANDARD; PRT; 220 AA.
AC Q60319;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0008.
GN MJ0008.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RC MEDLINE=96337999; PubMed=8688087;
RX

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii";
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: SOME, TO M.JANNASCHII MJ1311.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67460; AAB97997.1; -
DR TIGR; MJ0008; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 220 AA; 25836 MW; 955DF40FAB71B88 CRC64;

alignment_scores:
Quality: 31.00 Length: 8
Ratio: 3.875 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 50.000

alignment_block:
US-09-730-559B-45/rev x Y008_METJA ..
Align seg 1/1 to: Y008_METJA from: 1 to: 220

24 CGATCTATGATATATTAAGCCC 1
|||||:|||||:|||||:|||||:
175 LysAsnLeuAsnAsnLeuLysPro 182
seq_name: SwissProt_40:VG07_BPP22
seq_documentation_block:
ID VG07_BPP22 STANDARD; PRT; 229 AA.
AC Q01074;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA transfer protein gp7 precursor.
GN 7.
OS Bacteriophage P22.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae.
OX NCBI_TaxID=10754;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=92394890; PubMed=1522065;
RA Conlin C.A., Vimr E.R., Miller C.G.;
RT "Oligopeptidase A is required for normal phase P22 development.";
RL J. Bacteriol. 174:5869-5880(1992).
RN [2]
RN SEQUENCE FROM N.A.
RP Kropinski A.M.B., VanderByl C.S.;
RA "The completed sequence of genome of Salmonella phage P22.";
RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL [3]
RN SEQUENCE OF 225-229 FROM N.A.
RP MEDLINE=93219140; PubMed=8464750;
RA Adhikari P., Berget P.B.;
RT "Sequence of a DNA injection gene from Salmonella typhimurium phage
RT P22.";
RL Nucleic Acids Res. 21:1499-1499(1993).
RX [4]

```

```

RP SEQUENCE OF 21-31.
RX MEDLINE=91306435; PubMed=1853558;
RA Eppler K., Wyckoff E., Goates J., Parr R., Casjens S.;
RT "Nucleotide sequence of the bacteriophage P22 genes required for DNA
RT packaging.";
RL Virology 183:519-538(1991).
CC -!- FUNCTION: REQUIRED FOR EJECTION OF THE PHAGE DNA FROM THE PHAGE
CC PARTICLE AND INJECTION OF THAT DNA INTO THE HOST.
CC -!- SIMILARITY: STRONG, TO PHAGE APSE-1 P32.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M93985; AAA72115.1; -.
DR EMBL; AF217253; AAF75053.1; -.
DR EMBL; L07556; CAB23820.1; -.
DR PIR; C43330; C43330.
KW Late protein.
FT PROPEP 1 20 REMOVED IN MATURE FORM.
FT CHAIN 21 229 DNA TRANSFER PROTEIN GP7.
SQ SEQUENCE 229 AA; 23407 MW; 051A26642F185982 CRC64;

```

```

alignment_scores:
  Quality: 31.00      Length: 6
  Ratio: 5.167      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 83.333

```

alignment\_block:

US-09-730-559B-45/rev x VG07\_BPP22 ..

Align seg 1/1 to: VG07\_BPP22 from: 1 to: 229

```

18 ATGAATAATATTAAAGCCC 1
|||||
53 MetAsnLeuLysPro 58

```



OM of: US-09-730-559b-45 to: SPTRMBL\_19:\* out\_format : pfs  
Date: Sep 8, 2002 8:12 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=framet n2p.model -DEV=xlh  
-O/cn2\_1/USPTO\_SPOOL/US09730559/runat\_03092002\_150446\_3306/app\_query.fasta\_1.86  
-DB=SPTRMBL\_19 -QWTF=fastan -SUFFIX=rspt -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM\_ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09730559.ecgn1\_1.60 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT\_THREADS=1

Search information block:  
Query: US-09-730-559b-45  
Query length: 26  
Database: SPTRMBL\_19:\*  
Database sequences: 562222  
Database length: 172994929  
Search time (sec): 84.610000

## score\_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_invertebrate:018242	-	38.00	136.94	8.13	343	! 018242 caenorhabditis elegans
sp_invertebrate:016733	-	38.00	136.76	8.14	351	! 016733 caenorhabditis elegans
sp_plant:Q9LGM3	-	37.00	143.35	13.18	93	! Q9LGM3 oryza sativa (rice). hyp
sp_rodent:Q9P9T9	-	36.00	132.24	22.66	225	! Q9P9T9 mus musculus (mouse). 17
sp_human:Q96KZ8	+	36.00	132.17	22.66	227	! Q96KZ8 homo sapiens (human). si
sp_invertebrate:Q9W1Z5	+	35.00	118.17	39.34	788	! Q9W1Z5 drosophila melanogaster
sp_invertebrate:Q93636	+	35.00	111.62	40.22	1785	! Q93636 caenorhabditis elegans
sp_invertebrate:Q93637	+	35.00	111.00	40.30	1929	! Q93637 caenorhabditis elegans
sp_bacteriap:Q86710	+	34.00	125.86	63.47	182	! Q86710 aquifex aeolicus. hypoth
sp_bacteriap:Q99R86	+	34.00	123.07	64.07	258	! Q99R86 staphylococcus aureus (s
sp_bacteriap:Q9A9Z8	+	34.00	117.16	65.36	540	! Q9A9Z8 caulobacter crescentus.
sp_invertebrate:Q9U0J6	+	34.00	104.67	68.17	2567	! Q9U0J6 plasmodium falciparum
sp_human:Q9HAC7	-	33.00	114.87	109.05	434	! Q9HAC7 homo sapiens (human). hy
sp_plant:Q9XJ26	-	33.00	114.11	109.33	477	! Q9XJ26 nicotiana tabacum (commo
sp_bacteriap:Q87888	+	33.00	113.00	109.74	548	! Q87888 lawsonia intracellularis
sp_bacteriap:Q92PG2	+	33.00	112.96	109.76	551	! Q92PG2 rhizobium meliloti (sinc
sp_plant:Q941B6	-	33.00	112.84	109.80	559	! Q941B6 nicotiana tabacum (commo
sp_bacteriap:Q98DS0	-	33.00	112.44	109.95	588	! Q98DS0 rhizobium loti (mesorhiz
sp_invertebrate:Q9W0S9	+	33.00	105.80	112.44	1346	! Q9W0S9 drosophila melanogaster
sp_invertebrate:Q9NGP2	+	33.00	104.22	113.04	1640	! Q9NGP2 drosophila melanogaster
sp_bacteriap:Q9ETD0	+	32.00	120.06	177.43	137	! Q9ETD0 corynebacterium equi (i
sp_virus:Q87706	-	32.00	116.95	179.30	202	! P87706 fowlpox virus (fpv). ori
sp_invertebrate:Q9GR98	+	32.00	116.20	179.76	222	! Q9GR98 aphid gossypii (cotton
sp_human:Q9BVC5	-	32.00	115.84	179.97	232	! Q9BVC5 homo sapiens (human). hy
sp_virus:Q90B62	-	32.00	113.79	181.22	300	! Q9B62 yaba monkey tumor virus.
sp_bacteriap:Q981Y6	+	32.00	113.45	181.43	333	! Q981Y6 rhizobium loti (mesorhiz
sp_vertebrate:Q9FH55	+	32.00	113.29	181.53	319	! Q9FH55 gallus gallus (chicken).
sp_plant:Q9SRH3	+	32.00	113.20	181.59	323	! Q9SRH3 arabidopsis thaliana (mc
sp_bacteriap:Q9AEU1	+	32.00	112.93	181.75	334	! Q9AEU1 streptococcus gordonii.
sp_bacteriap:Q9PBK3	+	32.00	112.24	182.17	364	! Q9PBK3 xylella fastidiosa. abc
sp_bacteriap:Q9CBM5	+	32.00	112.02	182.31	374	! Q9CBM5 mycobacterium leprae. hy
sp_plant:Q9MAD5	+	32.00	111.94	182.36	378	! Q9MAD5 arabidopsis thaliana (mc
sp_invertebrate:Q9GZF0	-	32.00	111.73	182.49	388	! Q9GZF0 caenorhabditis elegans
sp_invertebrate:Q9XTH5	+	32.00	110.43	183.28	456	! Q9XTH5 caenorhabditis elegans
sp_invertebrate:Q9TVL3	+	32.00	109.87	183.63	489	! Q9TVL3 caenorhabditis elegans
sp_bacteriap:Q9CNX8	+	32.00	109.34	183.96	523	! Q9CNX8 pasteurella multocida. g
sp_bacteriap:Q9TPS2	+	32.00	109.24	184.02	529	! Q9TPS2 streptococcus pneumoniae
sp_plant:Q9LRF8	+	32.00	108.25	184.64	599	! Q9LRF8 arabidopsis thaliana (mc
sp_plant:Q93223	+	32.00	107.97	184.81	620	! Q93223 arabidopsis thaliana (mc
sp_plant:Q93Y36	+	32.00	107.77	184.94	636	! Q93Y36 arabidopsis thaliana (mc

sp\_fungi:Q9UTE5 + 32.00 107.73 184.96 639 ! Q9UTE5 schizosaccharomyces p  
sp\_plant:Q9M9M5 + 32.00 107.73 184.96 639 ! Q9M9M5 arabidopsis thaliana  
sp\_plant:Q9ZPW5 + 32.00 106.97 185.44 703 ! Q9ZPW5 arabidopsis thaliana  
sp\_invertebrate:Q93382 - 32.00 105.82 186.16 811 ! Q93382 caenorhabditis eleg  
sp\_bacteriap:Q9JTB6 + 32.00 104.56 186.95 950 ! Q9JTB6 neisseria meningitidi

seq\_name: sp\_invertebrate:018242

seq\_documentation\_block:

ID 018242 PRELIMINARY; PRT; 343 AA.  
AC 018242;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Y57G11C.18 PROTEIN.  
GN Y57G11C.18.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA McMurray A.A.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL: Z99281; CAB16519.1; -;  
SQ SEQUENCE 343 AA; 40942 MW; 617F6E8291BF77F8 CRC64;

alignment\_scores:

Quality: 38.00 Length: 8  
Ratio: 4.750 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:

US-09-730-559b-45/rev x 018242 ..  
Align seg 1/1 to: 018242 from: 1 to: 343

24 CGATCTATCAATAATATTAAGCCC 1  
:::|||||:::|||||  
222 LysSerMetAsnVallysPro 229

seq\_name: sp\_invertebrate:016733

seq\_documentation\_block:

ID 016733 PRELIMINARY; PRT; 351 AA.  
AC 016733;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE K09F6.1 PROTEIN.  
GN K09F6.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RC STRAIN-BRISTOL N2;  
RP Scheet P.;  
RT "The sequence of C. elegans cosmid K09P6.";  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RT Waterston R.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF016683; AAB66192.1; -;  
SQ SEQUENCE 351 AA; 41772 MW; 8C0BBE35CAA25B15 CRC64;

alignment\_scores:  
Quality: 38.00 Length: 8  
Ratio: 4.750 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:

US-09-730-559B-45/rev x O16733 ..  
Align seg 1/1 to: O16733 from: 1 to: 351

24 CGATCTATGAATAATATTAAAGCCC 1  
:::|||||:::|||||  
263 LysSerMetAsnVallysPro 270

seq\_name: sp\_plant:Q9LGM3

seq\_documentation\_block:

ID Q9LGM3 PRELIMINARY; PRT; 93 AA.  
AC Q9LGM3;  
DT 01-OCT-2000 (TREMREL. 15, Created)  
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)  
DT 01-MAR-2001 (TREMREL. 16, Last annotation update)  
DE HYPOTHETICAL PROTEIN.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
clone:P0041E1.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
clone:P0433F09.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF002521; BAA96761.1; -;  
DR EMBL; AF002539; BAB08200.1; -;  
SQ SEQUENCE 93 AA; 10539 MW; CDBA1B7C77BB9D25 CRC64;

alignment\_scores:  
Quality: 37.00 Length: 8  
Ratio: 4.625 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:

US-09-730-559B-45/rev x Q9LGM3 ..  
Align seg 1/1 to: Q9LGM3 from: 1 to: 93

24 CGATCTATGAATAATATTAAAGCCC 1  
|||||:::|||||:::|||||  
5 ArgSerIleAsnVallysPro 12

seq\_name: sp\_rodent:Q9D9T9

seq\_documentation\_block:

ID Q9D9T9 PRELIMINARY; PRT; 225 AA.  
AC Q9D9T9;  
DT 01-JUN-2001 (TREMREL. 17, Created)  
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)  
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)  
DE 1700029F09RIK PROTEIN.  
GN 1700029F09RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=TESTIS;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
DR EMBL; AK006487; BAB24612.1; -;  
DR MGD; MGI:1922873; 1700029F09RIK.  
DR InterPro; IPR000379; Est\_lip\_thioest\_actsite.  
SQ SEQUENCE 225 AA; 25240 MW; F713B5D975EF3249 CRC64;

alignment\_scores:  
Quality: 36.00 Length: 7  
Ratio: 5.143 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:

US-09-730-559B-45 x Q9D9T9 ..  
Align seg 1/1 to: Q9D9T9 from: 1 to: 225

1 GGCCTTAATATTATTCATAGA 21  
|||||:::|||||:::|||||  
68 GlyLeuAsnIleValHisarg 74

seq\_name: sp\_human:Q96KZ8

seq\_documentation\_block:

ID Q96KZ8 PRELIMINARY; PRT; 227 AA.  
AC Q96KZ8;



DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE SIMILAR TO RIKEN CDNA 1700029F09 GENE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Strausberg R.;  
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC015148; AAH15148.1; -;  
 SQ SEQUENCE 227 AA; 25585 MW; B37C06A38A91E774 CRC64;

alignment\_scores:  
 Quality: 36.00 Length: 7  
 Ratio: 5.143 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:

US-09-730-559B-45 x Q96KZ8 ..

Align seg 1/1 to: Q96KZ8 from: 1 to: 227

1 GGGCTTAATATTATTCATAGA 21  
 |||||  
 68 GlyLeuAsnIleValHisArg 74

seq\_name: sp\_invertebrate:Q9W1Z5

seq\_documentation\_block:

ID Q9W1Z5 PRELIMINARY; PRT; 788 AA.

AC Q9W1Z5;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE CG13528 PROTEIN.

GN CG13528.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agyaniani A., An H.-J., Andrews-Frankoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Fablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003458; AAF46906.1; -;  
 DR FlyBase; FBgn0034779; CG13528.  
 DR InterPro; IPR002110; ANK.  
 DR Pfam; PF00023; ank; 4.  
 DR SMART; SM00248; ANK; 3.  
 DR PROSITE; PS50088; ANK\_REPEAT; 3.  
 DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 2.  
 KW ANK repeat; Repeat.  
 SQ SEQUENCE 788 AA; 85586 MW; 5E03F8124B07474C CRC64;

alignment\_scores:  
 Quality: 35.00 Length: 8  
 Ratio: 4.375 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:

US-09-730-559B-45 x Q9W1Z5 ..

Align seg 1/1 to: Q9W1Z5 from: 1 to: 788

1 GGGCTTAATATTATTCATAGTCG 24  
 |||||  
 118 GlyLeuAsnIleHisLysAla 125

seq\_name: sp\_invertebrate:Q93636

seq\_documentation\_block:

ID Q93636 PRELIMINARY; PRT; 1785 AA.

AC Q93636;

DT 01-FEB-1997 (TREMBlrel. 02, Created)

DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE F29G6.3B PROTEIN.

GN F29G6.3B.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Harris B.R.;

RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

EX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for

RT investigating biology."

RL Science 282:2012-2018(1998).

DR EMBL; Z78543; CAB01754.1; -;

DR InterPro; IPE004019; YLP.

DR Pfam; PF02757; YLP; 8.

SQ SEQUENCE 1785 AA; 205049 MW; F92FFC5ABEC286CD CRC64;

alignment\_scores:  
 Quality: 35.00 Length: 8

```
Ratio: 4.375          Gaps: 0
Percent Similarity: 100.000    Percent Identity: 62.500

alignment_block:
US-09-730-559B-45 x Q93636  ..

Align seg 1/1 to: Q93636 from: 1 to: 1785

1 GGCCTTAATATTATTCATAGATCG 24
|||||:|||||:|||||:|||||
261 GlyValAsnValLeuHisArgSer 268

seq_name: sp_invertebrate:Q93637

seq_documentation_block:
ID Q93637 PRELIMINARY; PRT; 1929 AA.
AC Q93637;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F29G6.3A PROTEIN.
GN F29G6.3A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B.R.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; 278543; CAB01755.1; -.
SQ SEQUENCE 1929 AA; 220417 MW; C45166E4AAACE23B9 CRC64;

alignment_scores:
Quality: 35.00          Length: 8
Ratio: 4.375           Gaps: 0
Percent Similarity: 100.000    Percent Identity: 62.500

alignment_block:
US-09-730-559B-45 x Q93637  ..

Align seg 1/1 to: Q93637 from: 1 to: 1929

1 GGCCTTAATATTATTCATAGATCG 24
|||||:|||||:|||||:|||||
261 GlyValAsnValLeuHisArgSer 268

seq_name: sp_bacteriap:O66710

seq_documentation_block:
ID O66710 PRELIMINARY; PRT; 182 AA.
AC O66710;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 19.9 KDA PROTEIN.
GN AQ_389.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
```

```
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
DR EMBL; AE000888; AAC06687.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 182 AA; 19894 MW; DEF95FD6594D074 CRC64;

alignment_scores:
Quality: 34.00          Length: 7
Ratio: 4.857           Gaps: 0
Percent Similarity: 100.000    Percent Identity: 85.714

alignment_block:
US-09-730-559B-45 x O66710  ..

Align seg 1/1 to: O66710 from: 1 to: 182

1 GGCCTTAATATTATTCATAGA 21
|||||:|||||:|||||:|||||
43 GlyLeuAsnIleHisLys 49

seq_name: sp_bacteriap:Q99RE6

seq_documentation_block:
ID Q99RE6 PRELIMINARY; PRT; 258 AA.
AC Q99RE6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN SA2275 (HYPOTHETICAL PROTEIN SAV2487).
GN SA2275 OR SAV2487.
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879; 158878;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003137; BAB43578.1; -.
DR EMBL; AP003365; BAB58649.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 258 AA; 30455 MW; CCF5B0684B2D33BD CRC64;

alignment_scores:
Quality: 34.00          Length: 8
Ratio: 4.250           Gaps: 0
Percent Similarity: 100.000    Percent Identity: 62.500

alignment_block:
US-09-730-559B-45 x Q99RE6  ..

Align seg 1/1 to: Q99RE6 from: 1 to: 258

2 GGCCTTAATATTATTCATAGATCGA 25
|||||:|||||:|||||:|||||
94 GlyLeuValLeuTyrValAspArg 101
```

seq\_name: sp\_bacteriap:Q9A928

seq\_documentation\_block:

ID Q9A928 PRELIMINARY; PRT; 540 AA.  
 AC Q9A928;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOTHETICAL PROTEIN CC0813.  
 GN CC0813.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter  
 OX NCBI\_TaxID=69394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL; AF005758; AAK22798.1; -.  
 DR TIGR; CC0813; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 540 AA; 59648 MW; 72BC4542BEF99FD CRC64;

alignment\_scores:  
 Quality: 34.00 Length: 8  
 Ratio: 4.857 Gaps: 0  
 Percent Similarity: 87.500 Percent Identity: 87.500

alignment\_block:

US-09-730-559B-45 x Q9A928 ..

Align seg 1/1 to: Q9A928 from: 1 to: 540

2 GGCTTAATATATTCATAGATCGA 25  
 |||||  
 431 GlyLeulleuPheTyAspArg 438

seq\_name: sp\_invertebrate:Q9U0J6

seq\_documentation\_block:

ID Q9U0J6 PRELIMINARY; PRT; 2567 AA.  
 AC Q9U0J6;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOTHETICAL 312.5 KDA PROTEIN.  
 GN HAL4P2.26.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RA Devlin K., Bowman S., Churcher C., Harris B., Harris D., Lawson D.,  
 RA Quail M., Barrell B.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL035475; CAB62867.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 2567 AA; 312499 MW; F58B42994C211F47 CRC64;

alignment\_scores:  
 Quality: 34.00 Length: 8

Ratio: 4.250 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 62.500

alignment\_block:

US-09-730-559B-45 x Q9U0J6 ..

Align seg 1/1 to: Q9U0J6 from: 1 to: 2567

1 GGGCTTAATATATTCATAGATCG 24  
 |||||  
 932 GlyLeuAsnValIleHisLysAsn 939

seq\_name: sp\_human:Q9HAC7

seq\_documentation\_block:

ID Q9HAC7 PRELIMINARY; PRT; 434 AA.  
 AC Q9HAC7;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 DE HYPOTHETICAL 47.2 KDA PROTEIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=EMBRYO;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masubo Y., Oshima A.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK021870; BAB13922.1; -.  
 DR InterPro; IPR003673; CAIB-BAIF.  
 DR Pfam; PF02515; CAIB-BAIF; 1.  
 DR SEQUENCE 434 AA; 47186 MW; 7154D3E23618A883 CRC64;

alignment\_scores:  
 Quality: 33.00 Length: 6  
 Ratio: 5.500 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-730-559B-45/rev x Q9HAC7 ..

Align seg 1/1 to: Q9HAC7 from: 1 to: 434

18 ATCAATAATATTAGCCC 1  
 |||||  
 42 MetAsnIleLeysPro 47

seq\_name: sp\_plant:Q9XJ26

seq\_documentation\_block:

ID Q9XJ26 PRELIMINARY; PRT; 477 AA.  
 AC Q9XJ26;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE NTWRKY1.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. XANTHI;  
 RA Shinshi H., Yamamoto S., Suzuki K.;

RT "Analyses of an elicitor-responsive element and transcription factors  
RT in cultured tobacco cells."  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB022693; BAA82107.1; -  
DR InterPro; IPR003657; WRKY.  
DR Pfam; PF03106; WRKY; 2  
SQ SEQUENCE 477 AA; 53053 MW; 24B8617D8C59B7A9 CRC64;

alignment\_scores:  
Quality: 33.00 Length: 6  
Ratio: 5.500 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-730-559B-45/rev x Q9XJ26 ..  
Align seg 1/1 to: Q9XJ26 from: 1 to: 477

18 ATGATTAATATTAGCCC 1  
|||||  
456 MetAsnAsnIleLysPro 461

seq\_name: sp\_bacteria:O87888

seq\_documentation\_block:  
ID O87888 PRELIMINARY; PRT; 548 AA.  
AC O87888;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN).  
OS Lawsonia intracellularis.  
OG Plasmid pISI-2.  
OC Bacteria; Proteobacteria; delta subdivision; Lawsonia.  
OX NCBI\_TaxID=29546;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98386497; PubMed=9720028;  
RA Dale C.J.H., Moses E.K., Ong C.C., Morrow C.J., Reed M.B., Hasse D.,  
RA Strugnelli R.A.;  
RT "Identification and sequencing of the groE operon and flanking genes  
of Lawsonia intracellularis: use in phylogeny."  
RL Microbiology 144:0-0(0).  
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND  
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS  
CC CONDITIONS (BY SIMILARITY).  
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF  
CC 7 SUBUNITS (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.  
DR EMBL; U45241; AAC36500.1; -  
DR HSP; P06139; IGRU.  
DR InterPro; IPR001844; Chaperonins\_cpn60.  
DR InterPro; IPR002423; TCPL\_cpn60.  
DR Pfam; PF00118; cpn60\_TCPL; 1.  
DR PRINTS; PR00298; CHAPERONIN60.  
DR PRINTS; PR00304; TCOMPLEXTCP1.  
DR PROSITE; PS00296; CHAPERONIN\_CPN60; 1.  
KW ATP-binding; Chaperone; Plasmid.  
SQ SEQUENCE 548 AA; 58605 MW; 6388C431E663E498 CRC64;

alignment\_scores:  
Quality: 33.00 Length: 8  
Ratio: 4.125 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.500

alignment\_block:

US-09-730-559B-45 x O87888 ..  
Align seg 1/1 to: O87888 from: 1 to: 548

1 GGGCTTAATATTATCATGATCG 24

|||||  
439 GlyLeuAsnIleIleArgArgSer 446



